

**RANCANG BANGUN APLIKASI
QUESTION ANSWERING (QA) SYSTEM PADA
TERJEMAHAN AL QURAN MENGGUNAKAN *EPHYRA*
FRAMEWORK**

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Pada
Jurusan Teknik Informatika

Oleh:

NOVRENI ANGGRAINI
10851001750



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU
2013**

RANCANG BANGUN APLIKASI *QUESTION ANSWERING (QA)* SYSTEM PADA TERJEMAHAN AL QURAN MENGGUNAKAN *EPHYRA FRAMEWORK*

NOVRENI ANGGRAINI

10851001750

Tanggal Sidang: 02 Januari 2013

Periode Wisuda: Februari 2013

Jurusan Teknik Informatika

Fakultas Sains dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Al Quran merupakan tuntunan yang wajib diikuti oleh umat Islam. Permasalahan-permasalahan agama dalam kehidupan sehari-hari banyak termaktub pada Al Quran. Dalam Al Quran suatu permasalahan tidak hanya mengacu pada satu ayat ataupun satu surat saja sehingga dibutuhkan waktu yang lama dalam proses pencarian secara manual mengingat banyaknya jumlah ayat dan surat yang terkandung dalam Al Quran. Untuk itu dibutuhkan sebuah aplikasi yang mudah dalam mengenali dan mencari masalah yang dibutuhkan oleh pengguna sehingga sistem tersebut dapat menampilkan ayat-ayat Al Quran sebagai referensi. Berdasarkan hal tersebut pada penelitian ini, dibangun aplikasi QA *System* pada terjemahan Al Quran menggunakan Ephyra *Framework* dengan IR *system* yang digunakan adalah model ruang vektor yang dibangun dengan bahasa PHP. Dalam penelitian ini digunakan tiga kategori pertanyaan yang meliputi orang, tempat dan waktu dengan kata tanya yang digunakan adalah siapa, siapakah, kapan, kapankah, dimana, dimanakah, kemana, kemanakah, darimana, dan darimanakah. Hasilnya adalah secara keseluruhan aplikasi QA *System* ini memiliki nilai presisi sebesar 42,31% .

Kata kunci: *Ephyra Framework, Information Retrieval, Model Ruang Vektor, Question Answering System, Terjemahan Al Quran.*

DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL.....	iv
LEMBAR PERNYATAAN	v
LEMBAR PERSEMBAHAN	vi
ABSTRAK	vii
<i>ABSTRACT</i>	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xvii
DAFTAR RUMUS	xix
DAFTAR ISTILAH	xx
DAFTAR LAMPIRAN.....	xxi
DAFTAR SIMBOL.....	xxii
BAB I PENDAHULUAN	I-1
1.1. Latar Belakang	I-1
1.2. Rumusan Masalah.....	I-3
1.3. Batasan Masalah	I-3
1.4. Tujuan	I-4
1.5. Sistematika Penulisan	I-4
BAB II LANDASAN TEORI	II-1
2.1. Pengenalan Sistem Temu Kembali Informasi	II-1
2.2. Sistem Tanya Jawab (<i>Question Answering System</i>)	II-2
2.3. Arsitektur QA System	II-4
2.4. Ephyra Framework	II-6
2.4.1. Arsitektur Ephyra	II-6
2.4.2. <i>Pattern Learning and Matching</i>	II-12
2.4.3. Pola Bahasa Indonesia	II-14

2.5. Model Ruang Vektor	II-16
2.5.1. Pengindeksan Dokumen	II-17
2.5.2. Pembobotan Kata (<i>Term Wighting</i>).....	II-18
2.5.3. Ukuran Kesamaan (<i>Similarity Measure</i>)	II-19
2.6. Representasi Dokumen	II-19
2.7. Ekspresi Regular (<i>Regular Expression</i>)	II-20
2.7.1. Notasi Ekspresi Regular	II-20
2.8. Tata Bahasa Indonesia	II-22
2.8.1. Kalimat Tanya (Interogatif)	II-22
2.9. Evaluasi <i>Question Answering (QA) System</i>	II-23
BAB III METODOLOGI PENELITIAN	III-1
3.1. Alur Metodologi Penelitian	III-1
3.1.1. Pengumpulan Data.....	III-2
3.1.2. Analisa Sistem	III-2
3.1.3. Perancangan Sistem.....	III-3
3.1.4. Implementasi	III-3
3.1.5. Pengujian	III-4
3.1.6. Kesimpulan dan Saran.....	III-4
BAB IV ANALISA DAN PERANCANGAN.....	IV-1
4.1. Deskripsi Umum Sistem.....	IV-1
4.1.1. Analisa Kebutuhan Data.....	IV-3
4.1.2. Analisa Metode.....	IV-5
4.1.2.1. Pembentukan <i>Query (Query Formation)</i>	IV-6
4.1.2.2. Pencarian (<i>Search</i>)	IV-8
4.1.2.3. Seleksi Jawaban (<i>Answer Selection</i>)	IV-16
4.1.3. Analisa Fungsional Sistem	IV-23
4.1.4. Analisa Data Sistem.....	IV-26
4.2. Perancangan Sistem.....	IV-29
4.2.1. Perancangan Tabel.....	IV-29
4.2.2. Perancangan Menu	IV-32
4.2.3. Perancangan Antarmuka.....	IV-33

4.2.3.1. Tampilan Utama.....	IV-33
4.2.3.2. Tampilan Hasil Pencarian	IV-34
4.2.3.3. <i>Form Login</i> untuk <i>Administrator</i>	IV-35
4.2.3.4. Halaman Utama untuk <i>Administrator</i>	IV-35
BAB V IMPLEMENTASI DAN PENGUJIAN	V-1
5.1. Implementasi Sistem	V-1
5.1.1. Batasan Implementasi.....	V-1
5.1.2. Lingkungan Implementasi	V-1
5.1.3. Hasil Implementasi	V-2
5.2. Pengujian	V-9
5.2.1. Lingkungan Pengujian.....	V-9
5.2.2. Pengujian Sistem	V-10
5.2.2.1. Pengujian terhadap Pola Jawaban	V-10
5.2.2.2. Pengujian terhadap Aspek Presisi dan CWS...	V-17
5.2.2.3. Pengujian berdasarkan Interpretasi Pertanyaan	V-26
5.2.4. Kesimpulan Pengujian.....	V-30
BAB VI PENUTUP	VI-1
6.1. Kesimpulan.....	VI-1
6.2. Saran	VI-2
DAFTAR PUSTAKA	xxiii
LAMPIRAN	
RIWAYAT HIDUP	

DAFTAR GAMBAR

Gambar	Halaman
2.1 Contoh QA System pada yahoo answer	II-2
2.2 Contoh QA System pada ask.com	II-3
2.3 Contoh QA System pada answerbus.com.....	II-3
2.4 Arsitektur Umum dan Contoh Proses QA System.....	II-5
2.5 Arsitektur awal Ephyra Framework dengan pendekatan <i>pattern learning</i>	II-7
2.6 Arsitektur Ephyra Framework dengan penambahan teknik <i>answer type analysis</i> yang diterapkan pada TREC 2006.....	II-8
2.7 Arsitektur Ephyra Framework dengan penambahan teknik semantik ...	II-8
2.8 Contoh Tipe Jawaban Tempat.....	II-11
2.9 Ilustrasi <i>Similarity</i> 2 Dimensi.....	II-16
2.10 Bentuk Standar <i>Corpus</i>	II-19
3.1 Alur Metodologi Penelitian.....	III-1
4.1 Deskripsi Umum Sistem	IV-1
4.2 <i>Flowchart</i> Proses secara Keseluruhan	IV-2
4.3 Format Koleksi Dokumen Sistem	IV-9
4.4 <i>Context Diagram</i> Aplikasi QA System	IV-24
4.5 DFD Level 1 Aplikasi QA System.....	IV-25
4.6 ERD Aplikasi QA System	IV-27
4.7 Struktur Menu Aplikasi QA System.....	IV-33
4.8 Antarmuka <i>Form</i> Utama	IV-34
4.9 Antarmuka <i>Form</i> Hasil Pencarian.....	IV-34
4.10 Antarmuka <i>Form Login</i> untuk Administrator.....	IV-35
4.11 Antarmuka Halaman Utama untuk Administrator	IV-35
5.1 Menu Utama <i>Front End</i> Aplikasi QA System.....	V-2
5.2 Menu Utama <i>Back End</i> Aplikasi QA System.....	V-3
5.3 <i>Form</i> Tanya Jawab Aplikasi QA System	V-4

5.4	Daftar Jawaban Aplikasi QA System	V-4
5.5	Detil Jawaban Aplikasi QA System.....	V-5
5.6	Tampilan Terjemahan Al Quran Aplikasi QA System.....	V-5
5.7	Tampilan Korpus Terjemahan Al Quran Aplikasi QA System	V-6
5.8	Tampilan Halaman <i>Stoplist</i> Aplikasi QA System.....	V-6
5.9	Tampilan Halaman Kata Dasar Aplikasi QA System.....	V-7
5.10	Tampilan Halaman Pengindeksan Aplikasi QA System	V-7
5.11	Tampilan Halaman Segmentasi Aplikasi QA System	V-8
5.12	Tampilan Halaman Kamus Dasar Aplikasi QA System.....	V-9
E.1	DFD Level 2 Pengelolaan Data Master Aplikasi QA System	E-1
E.2	DFD Level 2 <i>Search</i> Aplikasi QA System	E-2
E.3	DFD Level 2 <i>Answer Selection</i> Aplikasi QA System.....	E-3
E.4	DFD Level 3 Pengelolaan Pengguna Aplikasi QA System.....	E-5
E.5	DFD Level 3 Pengelolaan Kata Tanya Aplikasi QA System	E-6
E.6	DFD Level 3 Pengelolaan Kata Dasar Aplikasi QA System.....	E-7
E.7	DFD Level 3 Pengelolaan <i>Stoplist</i> Aplikasi QA System.....	E-8
E.8	DFD Level 3 Pengelolaan Kamus Jawaban Aplikasi QA System.....	E-9
E.9	DFD Level 3 Pengelolaan Segmentasi Kalimat Aplikasi QA System ...	E-10
E.11	DFD Level 3 Pengelolaan Quran Aplikasi QA System.....	E-11
E.12	DFD Level 3 Pengindeksan Aplikasi QA System	E-12
F.1	Tampilan Menu Petunjuk pada Halaman <i>Front End</i>	F-1
F.2	Tampilan Menu Tentang QA System pada Halaman <i>Front End</i>	F-1
F.3	Tampilan <i>Login</i> pada Halaman <i>Back End</i>	F-2
F.4	Tampilan Halaman Pengguna pada Halaman <i>Back End</i>	F-2
F.5	Tampilan Tambah Data Pengguna pada Halaman <i>Back End</i>	F-2
F.6	Tampilan <i>Import</i> Data Pengguna pada Halaman <i>Back End</i>	F-3
F.7	Tampilan Pencarian Pengguna pada Halaman <i>Back End</i>	F-3
F.8	Tampilan Ubah Data Pengguna pada Halaman <i>Back End</i>	F-3
F.9	Tampilan Pencarian Data Kata Dasar pada Halaman <i>Back End</i>	F-4
F.10	Tampilan Tambah Data Kata Dasar pada Halaman <i>Back End</i>	F-4
F.11	Tampilan <i>Import</i> Data Kata Dasar pada Halaman <i>Back End</i>	F-4

F.12	Tampilan Pencarian <i>Stoplist</i> pada Halaman <i>Back End</i>	F-5
F.13	Tampilan Tambah Data <i>Stoplist</i> pada Halaman <i>Back End</i>	F-5
F.14	Tampilan <i>Import</i> Data <i>Stoplist</i> pada Halaman <i>Back End</i>	F-5
F.15	Tampilan Terjemahan Al Quran pada Halaman <i>Back End</i>	F-6
F.16	Tampilan Pencarian Terjemahan Al Quran pada Halaman <i>Back End</i> ...	F-6
F.17	Tampilan <i>Import</i> Data Terjemahan Al Quran pada Halaman <i>Back End</i>	F-6
F.18	Tampilan Pencarian Kamus Jawaban pada Halaman <i>Back End</i>	F-7
F.19	Tampilan Tambah Data Kamus Jawaban pada Halaman <i>Back End</i>	F-7
F.20	Tampilan Ubah Data Kamus Jawaban pada Halaman <i>Back End</i>	F-7
F.21	Tampilan Halaman Utama Kata Tanya pada Halaman <i>Back End</i>	F-8
F.22	Tampilan Pencarian Kata Tanya pada Halaman <i>Back End</i>	F-8
F.23	Tampilan Tambah Data Kata Tanya pada Halaman <i>Back End</i>	F-9
F.24	Tampilan Pencarian Data Segmentasi pada Halaman <i>Back End</i>	F-9

DAFTAR TABEL

Tabel	Halaman
2.1 Hasil Evaluasi Ephyra pada TREC 15 dan 16	II-6
2.2 Contoh Penerapan <i>Answer Type Filter</i>	II-12
2.3 Metakarakter pada Ekspresi Regular	II-20
2.4 Kelas Karakter pada Ekspresi Regular	II-21
4.1 Pola Pertanyaan (<i>Question Patterns</i>) untuk <i>Property</i> Orang,Tempat, dan Waktu	IV-7
4.2 Hasil <i>Parsing</i> pada Dokumen 1	IV-10
4.3 Hasil <i>Parsing</i> pada Dokumen 2	IV-10
4.4 Hasil <i>Parsing</i> pada Dokumen 3	IV-11
4.5 Hasil <i>Stopword Removing</i> pada Dokumen 1	IV-11
4.6 Hasil <i>Stopword Removing</i> pada Dokumen 2	IV-11
4.7 Hasil <i>Stopword Removing</i> pada Dokumen 3	IV-12
4.8 Hasil <i>Stemming</i> pada Dokumen 1	IV-12
4.9 Hasil <i>Stemming</i> pada Dokumen 2	IV-13
4.10 Hasil <i>Stemming</i> pada Dokumen 3	IV-13
4.11 Hasil Perhitungan TF IDF pada Pembobotan Kata.....	IV-14
4.12 Hasil Perangkingan Dokumen Berdasarkan Ukuran Kemiripan	IV-16
4.13 Hasil Segmentasi Kalimat pada Dokumen	IV-17
4.14 Ketentuan Tahapan Penyaringan Tipe Jawaban	IV-19
4.15 Hasil Penyaringan Tipe Jawaban	IV-20
4.16 Nilai Kandidat Jawaban	IV-20
4.17 Hasil Ekstraksi <TO> <CO>	IV-21
4.18 Pola Jawaban (<i>Answer Patterns</i>)	IV-21
4.19 Hasil Ekstraksi <PO>.....	IV-23
4.20 Kandidat Jawaban	IV-23
4.21 <i>Contect Diagram</i> Aplikasi QA System	IV-24
4.22 DFD Level 1 Aplikasi QA System	IV-25
4.23 ERD QA System.....	IV-27

4.24 Pengguna.....	IV-29
4.25 <i>Stoplist</i>	IV-29
4.26 Kata Dasar.....	IV-29
4.27 Quran.....	IV-30
4.28 Kemiripan	IV-30
4.29 Kamus Jawaban	IV-30
4.30 Kalimat.....	IV-30
4.31 Kata Tanya	IV-31
4.32 Jumlah <i>Keyword</i>	IV-31
4.33 Jawaban.....	IV-31
4.34 Kandidat <i>Answer Type</i>	IV-31
4.35 <i>Index</i>	IV-31
4.36 Pola Jawaban.....	IV-32
4.37 Bobot.....	IV-32
5.1 Deskripsi Menu Utama <i>Front End</i> Aplikasi QA System	V-2
5.2 Deskripsi Menu Utama <i>Back End</i> Aplikasi QA System	V-3
5.3 Pertanyaan yang Diajukan untuk Pola Jawaban	V-10
5.4 Hasil Pengujian Aplikasi terhadap Pola Jawaban	V-12
5.5 Hasil Pengujian Terhadap Aspek Presisi dan CWS pada Pengujian Pertama	V-19
5.6 Hasil Pengujian Terhadap Aspek Presisi dan CWS pada Pengujian Kedua.....	V-24
5.7 Nilai Presisi dan CWS	V-26
5.8 Hasil Pengujian Berdasarkan Interpretasi Pertanyaan	V-27

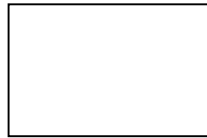
DAFTAR RUMUS

Rumus	Halaman
2.1. <i>Number of Keywords Filter</i>	II-10
2.2. Ekstraksi Pola Jawaban	II-13
2.3. Ekstraksi Pola Jawaban	II-13
2.4. <i>Confidence</i>	II-13
2.5. <i>Support</i>	II-13
2.6. <i>Term Weighting (TF)</i>	II-18
2.7. <i>Term Weighting (IDF)</i>	II-18
2.8. <i>Term Weighting (TF-IDF)</i>	II-18
2.9. <i>Similarity Measure</i>	II-19
2.10. <i>Precision of Correct</i>	II-24
2.11. <i>Precision of Incorrect</i>	II-24
2.12. <i>Confidence Wighted Score</i>	II-24

DAFTAR ISTILAH

<i>Information Retrieval</i>	suatu sistem yang digunakan untuk menemukan kembali (<i>retrieve</i>) informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis
<i>Inverse Document Frequency</i>	pengurangan dominasi <i>term</i> yang sering muncul di berbagai dokumen.
<i>Question Answering System</i>	kajian IR yang mengambil kutipan teks singkat dari dokumen yang dikembalikan sebagai jawaban dari <i>query</i> pertanyaan yang dimasukkan pengguna.
<i>Stemming</i>	proses pemetaan dan penguraian berbagai bentuk (<i>variants</i>) dari suatu kata menjadi bentuk kata dasarnya.
<i>Stopword Removing</i>	proses menyaring kata dengan mengambil kata-kata penting dan menghapus kata-kata yang tidak penting yang terdapat dalam <i>stoplist</i> .
<i>Term frequency</i>	frekuensi kemunculan <i>term</i> (t) pada dokumen (d).

DAFTAR SIMBOL



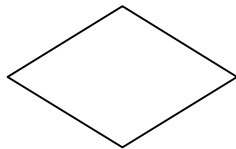
Proses pada *flowchart*



Start/Finish suatu proses pada *flowchart*



Alur/ langkah pada *flowchart* dan model data spasial



Kondisi/keadaan tindakan pada *flowchart* berupa Ya atau Tidak



Basis data pada *flowchart*

DAFTAR LAMPIRAN

Lampiran	Halaman
A. Daftar <i>Stopwords</i>	A-1
B. Dokumen yang Mengalami Penggabungan Ayat	B-1
C. Dokumen yang Memiliki Penanda Akhir Kalimat Berupa Tanda Seru dan Tanda Tanya.....	C-1
D. Daftar Kamus Jawaban	D-1
E. Rincian <i>Data Flow Diagram</i>	E-1
F. Implementasi dan Pengujian (Lanjutan)	F-1

BAB I

PENDAHULUAN

1.1. Latar Belakang

Question Answering (QA) System merupakan kajian sistem temu kembali informasi yang memiliki konsep dimana pengguna memasukkan sebuah *query* berupa pertanyaan dan sistem akan memberikan umpan balik berupa jawaban dari pertanyaan yang diajukan oleh pengguna secara langsung sehingga mempercepat proses penemuan masalah. *QA System* dapat dimanfaatkan dalam menyebarkan informasi pada berbagai aspek kehidupan tak terkecuali informasi pada aspek agama.

Seperti yang telah diketahui agama merupakan suatu aspek yang tidak dapat dipisahkan dalam kehidupan manusia. Tiap agama memiliki sebuah pedoman yang wajib diikuti oleh umat agama tersebut, begitu juga agama Islam. Dalam agama Islam, Al Quran dan hadits merupakan pedoman yang wajib diikuti umat Islam agar mendapatkan kehidupan yang sesuai dengan tuntunan agama Islam yang sebenarnya.

Al Quran sebagai pedoman utama berisi kumpulan firman-firman Allah dalam bahasa Arab. Untuk memudahkan umat Islam yang memiliki ragam bahasa memahami isinya, maka Al Quran telah diterbitkan dengan terjemahannya dalam berbagai bahasa termasuk bahasa Indonesia. Namun di sisi lain, dalam proses pencarian sebuah permasalahan pada Al Quran tidaklah mudah. Hal ini dikarenakan dalam Al Quran suatu permasalahan tidak mengacu pada satu ayat ataupun satu surat saja sehingga dibutuhkan waktu yang lama dalam proses pencarian secara manual mengingat banyaknya jumlah ayat dan surat yang terkandung dalam Al Quran. Oleh karena itu diperlukan media untuk memudahkan umat Islam dalam mengkaji Al Quran.

QA *System* bukanlah suatu ilmu baru dan telah banyak diteliti oleh peneliti-peneliti sebelumnya yang telah dipublikasikan diantaranya:

1. *A pattern learning approach to question answering within the Ephyra Framework* oleh N. Schlaefter, P. Giesermann, T.Schaff, dan A. Waibel (2006). Penelitian ini menjelaskan bagaimana cara kerja dari *pattern learning* pada *Ephyra framework*.
2. *The Ephyra QA System At TREC 2006* oleh N. Schlaefter, P. Giesermann dan G. Sautter (2006). Dalam penelitian ini N.Schlaefter dkk., membangun sebuah *Framework QA System* bahasa Inggris dengan menerapkan metode pencocokan pola (*pattern matching*) yang diberi nama *Ephyra* yang diterapkan pada TREC 2006 yang dapat diadaptasi untuk berbagai bahasa dan sistem terintegrasi lainnya. *Ephyra framework* mampu memberikan jawaban kepada pengguna berupa jawaban singkat (*factoid*).
3. *Pattern Based Approach in Indonesian Question Answering System* oleh Hapnes Toba dan Mirna Adriani (2009). Dalam penelitian ini dijelaskan bagaimana membuat sebuah pola (*pattern*) bahasa Indonesia dengan menggunakan pendekatan *pattern learning and matching* untuk QA *System* berbahasa Indonesia.

Ephyra merupakan sebuah *framework QA System* yang dikembangkan dengan bahasa Inggris yang telah dipublikasikan pada *Text Retrieval Conference (TREC)*. Kelebihan dari *Ephyra* adalah *Ephyra* mendukung integrasi dari algoritma dan sumber daya pengetahuan QA yang berbeda serta dapat diadaptasi ke bahasa lain. Dalam pengembangannya *Ephyra* menggunakan pendekatan *pattern learning and matching* untuk menjawab pertanyaan yang diajukan. Pendekatan ini lebih mudah diterapkan dalam menginterpretasikan pertanyaan dan mengekstrak jawaban.

Berdasarkan dari penjelasan di atas, maka pada Tugas Akhir ini penulis merancang bangun sebuah aplikasi *Question Answering (QA) System* berbasis web dengan menggunakan metode *Pattern learning and matching* bahasa Indonesia pada *Ephyra Framework* ke dalam sebuah korpus terjemahan Al Quran. Keluaran yang

dihasilkan nantinya berupa potongan kalimat sebagai jawaban yang diurutkan berdasarkan *score* yang didapatkan dari hasil proses pencocokan jawaban.

1.2. Rumusan Masalah

Berdasarkan penjelasan yang telah dijelaskan dibagian latar belakang di atas, maka dapat ditarik sebuah rumusan masalah yang akan dijelaskan lebih lanjut pada penelitian Tugas Akhir ini, yaitu "Bagaimana merancang bangun aplikasi *Question Answering (QA) System* berbasis web pada terjemahan Al Quran bahasa Indonesia dengan mengadaptasi Ephyra *framework*".

1.3. Batasan Masalah

Agar tidak terjadi kesalahan persepsi dalam penelitian tugas akhir ini, maka berikut dijelaskan beberapa hal yang menjadi batasan masalah yaitu:

1. Dalam pembuatan aplikasi *QA System* ini menggunakan metode pencocokan pola (*pattern matching*) bahasa Indonesia.
2. Kalimat pertanyaan yang menjadi masukan hanya terbatas pada pertanyaan seputar lokasi, waktu, dan orang dengan kata tanya yang digunakan adalah siapa, siapakah, kapan, kapankah, dimana, dimanakah, kemana, kemanakah, darimana, dan darimanakah.
3. Dokumen yang digunakan adalah terjemahan Al Quran dalam bahasa Indonesia (terjemahan DEPAG 1989). Dokumen memiliki format teks (*.txt) dan tidak menyertakan informasi yang terdapat pada catatan kaki.
4. Dalam pencarian dokumen, model sistem temu kembali informasi (*Information Retrieval*) yang digunakan adalah model ruang vektor.
5. Dalam proses pengindeksan, algoritma *stemming* yang digunakan adalah algoritma Nazief dan Adriani.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai pada penulisan tugas akhir ini adalah merancang bangun sebuah aplikasi QA *System* berbasis web pada terjemahan Al Quran bahasa Indonesia sebagai korpus dengan menggunakan Ephyra *framework*.

1.5. Sistematika Penulisan

Dalam penyusunan laporan Tugas Akhir ini, sistematika penulisan yang penulis gunakan adalah sebagai berikut :

1. Bab I Pendahuluan

Bab I ini merupakan bagian yang akan menguraikan hal-hal seperti: latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan laporan tugas akhir.

2. Bab II Landasan Teori

Bab ini berisi tentang teori-teori tentang *Question Answering (QA) System* dan metode-metode yang digunakan pada perancangan aplikasi QA *System* yang terdiri dari pengenalan sistem temu kembali informasi, sistem tanya jawab (*Question Answering System*), arsitektur QA *System*, Ephyra *Framework*, model ruang vektor, representasi dokumen, ekspresi regular, tata bahasa Indonesia, evaluasi *Question Answering (QA) System*.

3. Bab III Metodologi Penelitian

Bab ini berisi tentang cara-cara atau hal-hal yang dilakukan dalam menyelesaikan kasus tugas akhir ini yang mencakup alur metodologi penelitian, meliputi: pengumpulan data, analisa sistem, perancangan sistem, implementasi, pengujian serta kesimpulan dan saran.

4. Bab IV Analisa dan Perancangan

Bab ini berisi tentang analisa dari penelitian yang dilakukan dalam tugas akhir ini sekaligus menerangkan perancangan rancang bangun aplikasi *Question Answering (QA) System* yang dibangun. Pembahasan mengenai

analisa dan perancangan mencakup deskripsi umum sistem, analisa kebutuhan data, analisa metode, analisa fungsional sistem, analisa data sistem, perancangan tabel, perancangan menu, dan perancangan antarmuka.

5. Bab V Implementasi dan Pengujian

Bab ini berisi tentang langkah-langkah pembangunan rancang bangun aplikasi *Question Answering (QA) System* dan menguji hasil dari rancangan yang telah dibangun. Bab ini mencakup pembahasan mengenai batasan implementasi, lingkungan implementasi dan hasil implementasi. Pengujian mencakup lingkungan pengujian, pengujian terhadap pola jawaban, presisi, *Confidence Weighted Score (CWS)*, interpretasi pertanyaan dan kesimpulan pengujian.

6. Bab VI Penutup

Bab ini berisi kesimpulan dan saran mengenai hasil analisa, perancangan, hasil implementasi dan hasil pengujian yang telah dilakukan terhadap rancang bangun aplikasi *Question Answering (QA) System* yang telah dibangun.

BAB II

LANDASAN TEORI

2.1. Pengenalan Sistem Temu Kembali Informasi

Sistem temu kembali informasi atau *Information Retrieval System* merupakan salah satu tipe sistem informasi yang menjadi bagian dari *computer science* yang berhubungan dengan pengambilan informasi dari dokumen-dokumen yang didasarkan pada isi dan konteks dari dokumen-dokumen itu sendiri. Sistem temu kembali informasi telah banyak diterapkan pada mesin pencarian seperti Google, yahoo, Altavista dan lain-lain.

Secara harfiah sistem temu kembali informasi dapat diartikan sebagai suatu sistem yang digunakan untuk menemukan kembali (*retrieve*) informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. (Bunyamin, 2008).

Dalam sistem temu kembali informasi, banyak obyek yang menjadi kajian sistem temu kembali informasi, diantaranya:

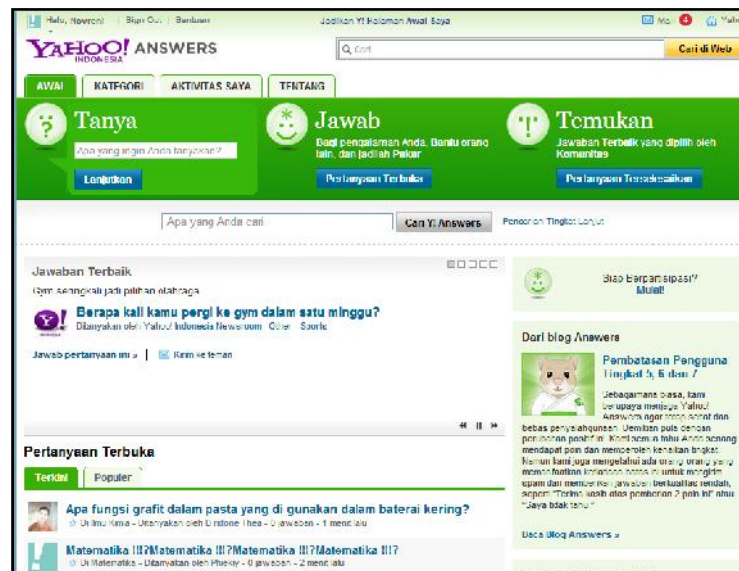
1. *Question answering IR system*, kajian IR yang mengambil kutipan teks singkat dari dokumen yang dikembalikan sebagai jawaban dari *query* pertanyaan yang dimasukkan pengguna.
2. *Cross-language retrieval*, kajian IR yang menggunakan *query* dalam satu bahasa misalnya bahasa Inggris dan menemukan dokumen dalam bahasa lain misalnya bahasa Indonesia.
3. *Speech retrieval*, kajian IR yang menggunakan suara sebagai data yang diolah.
4. *Music retrieval*, kajian IR yang menemukan potongan melodi ketika pengguna memasukkan catatan dari tema musik.

5. *Image retrieval*, kajian IR yang menemukan gambar pada tema atau gambar yang mengandung suatu bentuk atau warna.

2.2. Sistem Tanya Jawab (*Question Answering System*)

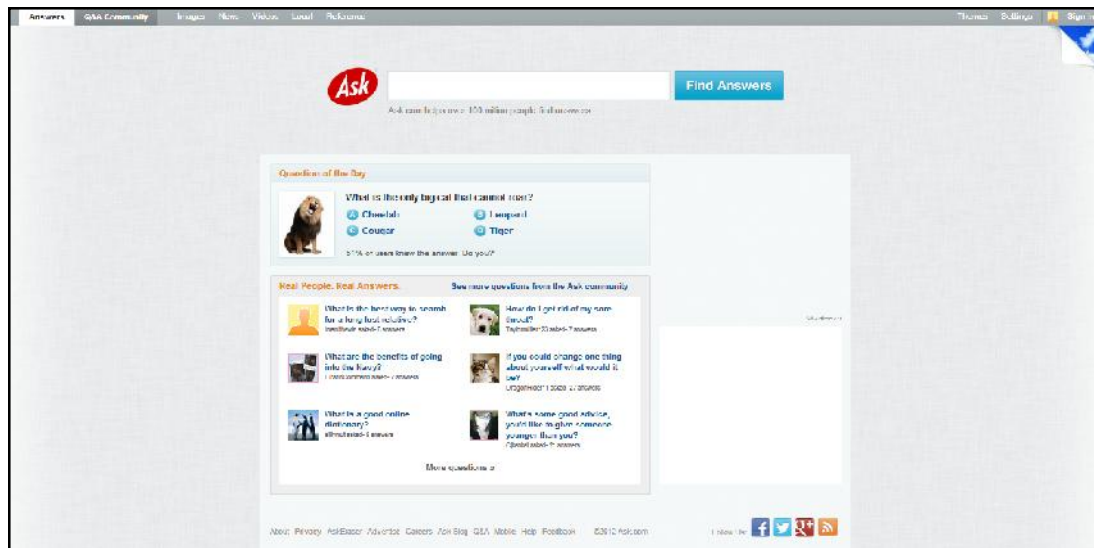
Question Answering (QA) System merupakan sebuah sistem yang memungkinkan pengguna menyatakan kebutuhan informasinya dalam bentuk yang lebih spesifik dan alami, yaitu dalam bentuk pertanyaan bahasa alami (*natural language question*), dan tidak mengembalikan daftar dokumen yang harus disaring oleh pengguna untuk menentukan apakah dokumen-dokumen tersebut mengandung jawaban atas pertanyaan, tetapi mengembalikan kutipan teks singkat atau bahkan frase sebagai jawaban. (Gunawan, dkk, 2006).

QA *System* memungkinkan *user* memasukkan pertanyaan dengan bahasa alami, yaitu bahasa yang digunakan manusia untuk berkomunikasi dalam kehidupan sehari-hari, seperti bahasa Inggris, Indonesia, dan lain-lain. Contoh aplikasi QA *System* seperti yahoo answer, ask.com, answerbus.com, dan lain sebagainya. (Lihat Gambar 2.1, 2.2 , dan 2.3).

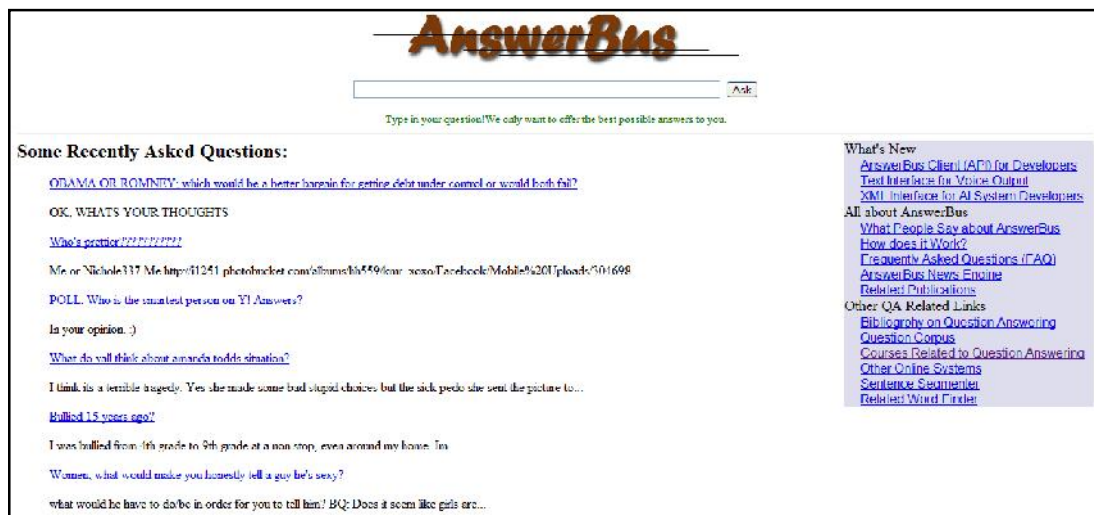


Gambar 2.1. Contoh QA *System* pada yahoo answer

(Sumber: id.answers.yahoo.com)



Gambar 2.2. Contoh QA System pada ask.com
(Sumber: www.ask.com)



Gambar 2.3. Contoh QA System pada answerbus.com
(Sumber: www.answerbus.com)

Dalam pengembangan sebuah QA System dibutuhkan pengetahuan dari berbagai disiplin ilmu seperti *Natural Language Processing*, Sistem Temu Kembali Informasi, dan *Information Extraction*. Saat ini, ketertarikan akan pengembangan QA

System banyak dipengaruhi oleh *Text Retrieval Conference* (TREC) dan *World Wide Web* (WWW).

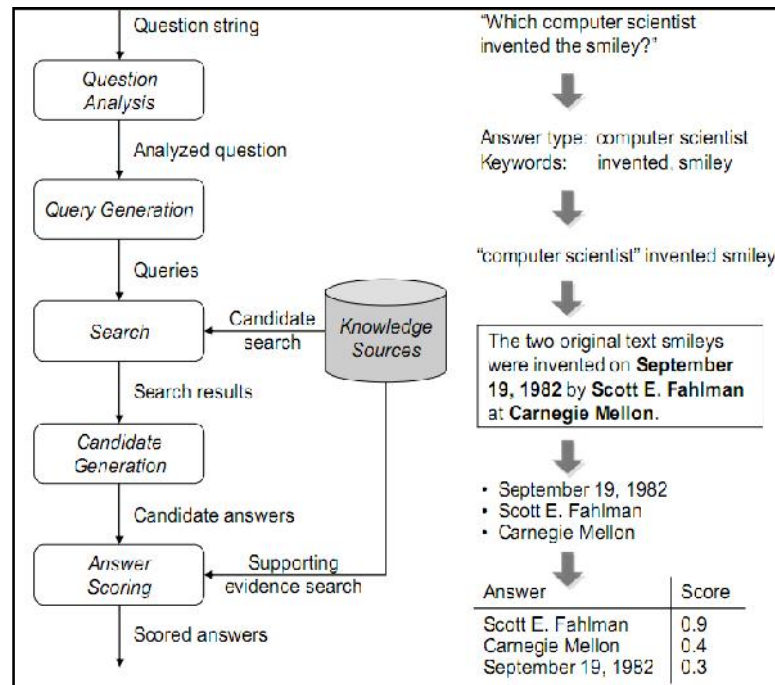
2.3. Arsitektur QA System

Dalam pembangunan QA *System* banyak arsitektur QA *System* yang dapat diadopsi. Namun secara umum QA *System* didasarkan pada komponen inti yaitu: analisa pertanyaan (*question analysis*), pembentukan *query* (*query generation*), pencarian (*search*), pembentukan kandidat jawaban dan penilaian jawaban (*candidate answer generation and answer scoring*). (Schlaefler, 2009-2011)

Analisa pertanyaan (*question analysis*) merupakan tahap awal yang dilakukan dari sebuah QA *System*. Komponen analisa pertanyaan berasal dari informasi sintaksis atau semantik dari pertanyaan dengan menggunakan teknik-teknik seperti klasifikasi tipe jawaban, penguraian (*parsing*) sintaksis dan semantik, dan pengenalan entitas bernama (*named entity recognition*). Pada tahap pembentukan *query* (*query generation*), informasi diubah menjadi sebuah set *query* pencarian yang akan diteruskan ke komponen pencarian untuk menemu kembalikan konten yang relevan dari kumpulan sumber pengetahuan.

Hasil pencarian akan diproses oleh *candidate generation* yang mengekstrak kandidat dari jawaban.

Selanjutnya, *answer scoring* akan memperkirakan nilai kepercayaan dari kandidat jawaban, kemudian merangkingnya, dan menggabungkan kandidat jawaban yang sama. Arsitektur QA *System* tersebut dapat dilihat pada Gambar 2.4.



Gambar 2.4. Arsitektur Umum dan Contoh Proses QA System

(Sumber: Schlaefer, 2009-2011)

Pada Gambar 2.4 di atas, di sisi kiri digambarkan bagaimana arsitektur dari QA System pada umumnya yang terdiri dari *question analysis*, *query generation*, *search*, *candidate generation*, dan *answer scoring*. Di sisi kanan, digambarkan bagaimana sebuah pertanyaan diproses dalam QA System. Misalnya, sebuah masukan pertanyaan berupa "Which computer scientist invented the smiley?". Pada tahap *question analysis*, tipe jawaban yang terbentuk dari pertanyaan tersebut berupa *computer scientist* dengan dua buah kata kunci tambahan yaitu: "invented" dan "smiley". Dari tipe jawaban dan kata kunci tambahan tersebut, *query generation* akan membentuk sebuah *query* mesin pencarian. Setelah itu pada tahapan *search*, komponen tersebut akan menemukan kembali bagian dari korpus teks (misalnya, web) seperti yang ditunjukkan dalam contoh. Selanjutnya entitas yang ada dapat diekstraksi sebagai kandidat jawaban di tahapan *candidate generation*. Dan terakhir, *answering scoring* akan memperkirakan nilai kepercayaan dari masing-masing

kandidat dengan menggunakan fitur seperti *retrieval rank*, jumlah kemunculan kandidat pada hasil pencarian, dan kesesuaian dengan tipe jawaban yang diprediksi. Kandidat yang memperoleh nilai tertinggi akan dikembalikan sebagai jawaban yang paling mungkin yaitu Scott E. Fahlman.

2.4. Ephyra Framework

Ephyra merupakan sebuah modular *framework* untuk membangun *Question Answering* (QA) *System* yang mendukung integrasi dari algoritma dan sumber daya pengetahuan QA yang berbeda. Ephyra dapat dengan mudah diadaptasi dengan bahasa lain selain bahasa Inggris. Situs resmi Ephyra www.ephyra.info.

Ephyra dirilis oleh Carnegie Mellon University pada Februari 2008 sebagai *framework* berbasis *open source*. Sebelum dirilis, Ephyra telah berkompetisi di beberapa QA TREC. Presisi dan *recall* yang dihasilkan Ephyra di tahun 2007 ditempatkan pada peringkat tiga teratas dalam hal akurasi *factoid*. Tabel 2.1 berikut menunjukkan hasil evaluasi Ephyra pada TREC 15 dan 16.

Tabel 2.1 Hasil Evaluasi Ephyra pada TREC 15 dan 16

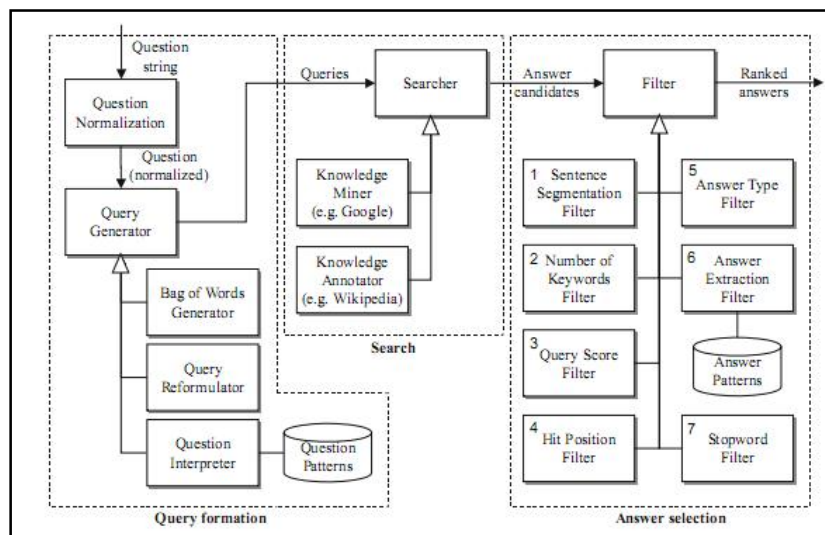
	<i>Factoid Accuracy</i>	<i>List F(1)</i>	<i>Definition F(3)</i>	<i>Overall Per-series</i>
TREC 15 Ephyra	0,196	0,096	0,150	0,143
Median	0,186	0,087	0,125	0,134
TREC 16 Ephyra	0,206	0,140	0,189	0,181
Median	0,131	0,085	0,118	0,108

(Sumber : Schlaefer, 2009-2011)

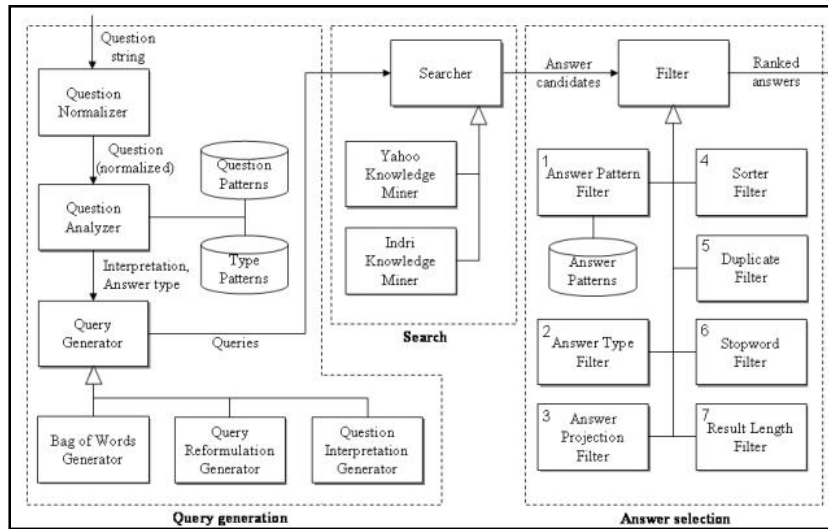
2.4.1. Arsitektur Ephyra

Sejak dirilisnya Ephyra hingga sekarang, Ephyra mengalami berbagai perkembangan. Dalam kemunculan awal tahun 2006, Ephyra memiliki arsitektur yang terdiri dari tiga tahapan utama, yaitu *query formation*, *search*, dan *answer selection*. (Lihat Gambar 2.5). Masih di tahun yang sama 2006, Ephyra mengalami modifikasi dengan penambahan *answer type analysis* yang terdiri dari tiga tahapan

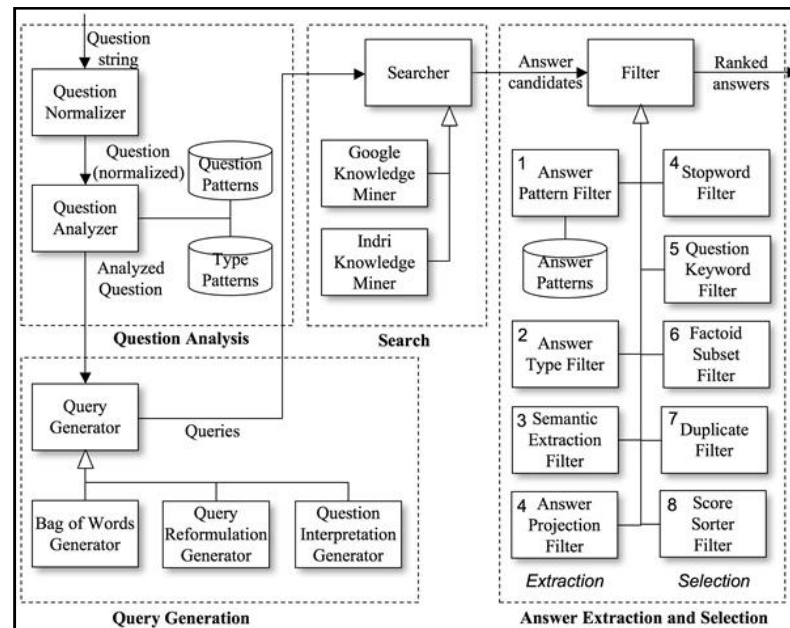
utama yaitu *query generation*, *search*, dan *answer selection*. (Lihat Gambar 2.6). Pada perkembangan berikutnya Ephyra mengalami modifikasi dengan penambahan teknik semantik yang terdiri dari empat tahapan utama yaitu *question analysis*, *query generation*, *search*, dan *answer extraction and selection*. (Lihat Gambar 2.7). Arsitektur-arsitektur ini memudahkan untuk bereksperimen dengan berbagai pengaturan dan memungkinkan untuk diintegrasikan dengan multi-teknik QA System ke dalam sebuah sistem tunggal.



Gambar 2.5. Arsitektur awal Ephyra Framework
dengan pendekatan *pattern learning*
(Sumber: Schlaefer,dkk, 2006)



Gambar 2.6. Arsitektur Ephra Framework dengan penambahan teknik *answer type analysis* yang diterapkan pada TREC 2006
(Sumber: Schlaefer,dkk, 2006)



Gambar 2.7. Arsitektur Ephra Framework dengan penambahan teknik *semantik*
(Sumber: Schlaefer,dkk, 2006)

Seperti yang telah dijelaskan di atas, saat ini konfigurasi Ephyra pada tahapan *answer extraction* dikombinasikan dengan tiga pendekatan (Schlaefter, 2009-2011), yaitu:

1. *Answer type analysis*, pendekatan ini dilakukan dengan menentukan jenis jawaban yang diharapkan dari sebuah pertanyaan dengan memetakan tipe hirarki dari *named entity (NE)* menggunakan *NE tagger* yang sesuai.
2. *Pattern learning and matching*, pendekatan ini digunakan untuk mengkategorikan pertanyaan dengan mencocokkan pola dari masing-masing jenis pertanyaan.
3. *Semantic parsing*, pendekatan ini menggunakan dua buah teknik semantik yaitu teknik parsing semantik yang digunakan untuk mengidentifikasi frase dalam sumber pengetahuan yang mirip dengan pertanyaan pada tingkat semantik dan ontologi untuk mengekstrak istilah kunci dari pertanyaan dan dokumen korpus.

Arsitektur pada Gambar 2.5 dengan pendekatan *pattern learning* di atas dapat diuraikan sebagai berikut:

1) *Query Formation*

Query formation adalah tahapan untuk membentuk sebuah *query* yang akan digunakan pada tahap *search*. Tahapan pada *query formation* terdiri dari dua tahapan utama yaitu tahapan *question normalization* dan tahapan *query generator*.

a. *Question normalization*

Pada tahapan ini kalimat pertanyaan akan dinormalisasi, normalisasi yang dilakukan dapat berupa menghilangkan tanda baca, menghilangkan token-token yang tidak perlu (*stopword removing*), menjadikan verba dan nomina menjadi kata dasar (*stemming*), dan sebagainya.

b. *Query generator*

Tahapan ini bertujuan untuk mengubah kalimat pertanyaan menjadi *query*. *Query generator* memiliki tiga komponen utama yaitu *bag of words generator*, *query reformulation*, dan *question interpreter*. Pada tahapan ini diterapkan *pattern matching* untuk pola pertanyaan yang dijelaskan pada subbab 2.4.2.

2) *Search*

Tahapan ini merupakan bagian dari sistem temu kembali informasi yang digunakan untuk mencari dokumen-dokumen yang relevan dengan *query* yang dihasilkan pada tahapan sebelumnya. Pada tahapan *search* dalam Tugas Akhir ini menggunakan model ruang vektor (Lihat subbab 2.5) untuk mencari dokumen-dokumen yang relevan.

3) *Answer Selection*

Tahapan ini adalah tahapan yang bertujuan untuk mendapatkan jawaban dari pertanyaan yang diajukan. *Answer selection* memiliki tahapan-tahapan sebagai berikut:

1. *Sentence segmentation filter*

Tahapan ini berfungsi untuk memecah dokumen menjadi kalimat-kalimat.

2. *Number of keywords filter*

Pada tahapan ini *QA System* akan menentukan kalimat-kalimat yang akan menjadi kandidat jawaban dari pertanyaan yang telah dimasukkan. Kalimat-kalimat tersebut akan menjadi kandidat jawaban jika memenuhi persamaan 2.1 (Zheng, 2002) berikut:

$$q \geq \lfloor \sqrt{Q-1} \rfloor + 1 \quad \dots\dots\dots(2.1)$$

dimana:

- q : jumlah *keyword query* yang ditemukan pada kalimat
- Q : jumlah *keyword* penyusun *query*

Selanjutnya, kalimat-kalimat yang memenuhi persamaan 2.1 akan diteruskan pada proses berikutnya dan kalimat-kalimat yang tidak memenuhi syarat akan dibuang dari daftar kandidat jawaban.

3. *Query score filter*

Sebuah filter yang menentukan kandidat jawaban dengan menghitung skor dari *query*.

4. *Hit position filter*

Sebuah filter yang menentukan kandidat jawaban dengan melihat posisi dokumen yang dikembalikan oleh mesin pencari.

5. *Answer type filter*

Answer type salah satu filter yang sering digunakan. Pada tahapan ini kandidat-kandidat jawaban akan diberi sebuah nilai estimasi. Kandidat-kandidat jawaban yang mempunyai tipe yang sama maka nilai estimasinya akan diakumulasikan. Contoh dari tipe jawaban seperti tempat, waktu, orang, dan lain sebagainya. Pada Gambar 2.8 dijelaskan contoh aturan yang dipakai untuk pertanyaan dengan tipe tempat dan Tabel 2.2 menjelaskan cara penerapannya.

```
AnswerType:
PLACE
QuestionPatterns:
(?i)(.*, )?where.*
1.0
(?i).*(what|which|name).*(capital|city|continent|country|
    mountain|place|river|state).*
1.0
AnswerPatterns:
[NE:res/namedentities/PLACE]
0.8
.*(in|on|at) [A-Z].*
0.8
.* [A-Z].*
0.3
```

Gambar 2.8. Contoh Tipe Jawaban Tempat (*Sumber: Schlaefel, 2005*)

Tabel 2.2 Contoh Penerapan *Answer Type Filter*

Pertanyaan	Where is Pittsburgh?
Tipe	Tempat
Nilai pertanyaan	1,0
Kandidat jawaban	Pittsburgh is in Pennsylvania
Nilai jawaban	0,8
Nilai	$1,0 \times 0,8 = 0,8$

6. *Answer extraction filter*

Answer extraction filter merupakan *filter* yang berfungsi untuk mengekstrak kandidat jawaban yang terdapat pada kalimat-kalimat dokumen. Pada tahapan ini digunakan *pattern matching* pola jawaban yang dijelaskan pada subbab 2.4.2.

7. *Stopword filter*

Filter ini berfungsi untuk membuang kandidat-kandidat jawaban yang termasuk ke dalam *stopword*.

2.4.2. *Pattern Learning and Matching*

Ephyra menggunakan pendekatan *pattern learning* dalam mengkategorikan pertanyaan. Kategorisasi tersebut dilakukan dengan mencocokkan pola masing-masing jenis pertanyaan, dan mencoba untuk mengkategorikan pertanyaan berdasarkan posisi kata tanya, dan kata kunci dari pertanyaan. Ephyra dapat mempelajari pola teks untuk ekstraksi jawaban. Ephyra dapat belajar tanya jawab dan menggunakan sistem pencarian umum yaitu mesin penelusur web atau berbagai sistem IR lainnya untuk mengambil teks dokumen. Ephyra menggunakan dua jenis dalam pendekatan *pattern learning* (Schlaefel, 2006) yaitu:

a. Pola Pertanyaan

Tujuan dari pola pertanyaan adalah untuk menafsirkan pertanyaan dan mengubahnya menjadi *query*. Strategi penting dalam pendekatan ini adalah menyederhanakan pertanyaan menjadi tiga komponen yang terdiri dari properti (P), target (T), dan konteks (C). *Template* dari pertanyaan perlu dikembangkan secara

manual sesuai dengan interogatif dari berbagai kalimat yang independen untuk setiap bahasa alami.

Ephyra menginterpretasikan pertanyaan dengan pencocokan yang sekuensial dari semua pola pertanyaan untuk setiap properti. Jika sesuai dengan pola, Ephyra akan mengekstrak objek target dan objek konteks. Oleh karena itu, sangat memungkinkan bahwa sebuah pertanyaan memiliki lebih dari satu interpretasi jika pertanyaan tersebut tidak cukup jelas.

b. Pola Jawaban

Tujuan dari pola ini adalah untuk mengekstrak kandidat jawaban dari potongan dokumen yang relevan dan untuk perangkingan dari kandidat jawaban tersebut. Sama halnya dengan pola pertanyaan, pola jawaban memiliki format yang terdiri dari: *tag* target <T>, *tag* konteks <C>, dan *tag* properti <P>. Selama tahap ekstraksi jawaban, Ephyra menggantikan semua kejadian dari objek target atau objek konteks pada potongan teks. Setiap potongan teks sesuai dengan pola, bagian dari teks yang terkait ke *tag* properti diekstrak. Untuk ekstraksi jawaban, Ephyra menerapkan dua macam ekspresi regular (*regular expression*) (Zhang, 2002) yaitu:

$$\backslash B < T > \backslash B (. * ?) \backslash B < P > \backslash B \backslash s * (\backslash W | w +) \dots\dots\dots(2.2)$$

$$(\backslash W | \backslash w +) \backslash s * \backslash B < P > \backslash B (. * ?) \backslash B < T > \backslash B \dots\dots\dots(2.3)$$

Setiap pola jawaban yang digunakan untuk mengekstrak jawaban ditentukan dari nilai *confidence* dan *support*-nya. Kandidat jawaban dengan nilai *confidence* tertinggi akan dianggap sebagai kandidat jawaban terbaik, dimana rumus *confidence* dan *support* dapat dilihat pada persamaan 2.4. dan 2.5 berikut:

$$confidence_a = \frac{correct_a}{correct_a + incorrect_a} \dots\dots\dots(2.4)$$

dimana:

$correct_a$: jawaban yang benar

$incorrect_a$: jawaban yang salah

$$support_a = \frac{correct_a}{snippets_p} \dots\dots\dots(2.5)$$

dimana:

$correct_a$: jawaban yang benar

$snippets_p$: total potongan jawaban untuk setiap *property*

2.4.3. Pola Bahasa Indonesia

Dalam membangun QA *system* dengan pendekatan pola bahasa Indonesia, ada beberapa hal yang harus diperhatikan, yaitu:

a. Kategorisasi pertanyaan

Seperti halnya bahasa Inggris, *pattern learning* pada pola bahasa Indonesia juga memiliki dua tahapan utama dalam mengkategorikan pertanyaan. Pertama, mempelajari pola pertanyaan dari *template* pertanyaan sesuai dengan masing-masing tipe pertanyaan. Kedua, mempelajari pola jawaban dari pasangan tanya jawab. (Toba,dkk, 2009)

Contoh pertanyaan “Dimana letak negara Indonesia?”

Jika pertanyaan tersebut diinterpretasikan akan menjadi:

Properti: tempat

Target: negara

Konteks: indonesia

Pertanyaan tersebut meminta properti dari objek target “indonesia” yang merupakan sebuah tempat. Objek konteks “Indonesia” mempersempit pencarian ke tempat tertentu. Selanjutnya ketiga objek tersebut digunakan untuk mengekstrak jawaban yang sesuai. Potongan dokumen yang mengandung unsur target <T> dan konteks <C> akan di-replace dengan tag <T> dan <C>, sehingga menghasilkan jawaban. Contoh terdapat potongan dokumen berikut:

“Negara Indonesia terletak di antara samudera hindia dan samudera pasifik. ” maka akan diubah menjadi “<T> <C> terletak di antara samudera hindia dan samudera pasifik”. Pola “<T> <C> terletak di antara samudera hindia dan samudera pasifik” digunakan untuk mengekstrak *property* samudera hindia dan samudera pasifik.

b. Membangun pola pertanyaan

Dalam membuat pola pertanyaan, perlu diperhatikan tipe dari pertanyaan tersebut agar jawaban yang dihasilkan sesuai dengan pertanyaan. Tipe pertanyaan menurut CLEF (Forner,dkk, 2008) terdiri dari 8 tipe, yaitu:

1. Orang
2. Waktu
3. Tempat
4. Organisasi
5. Ukuran
6. Angka
7. Obyek
8. Lain-lain

Pola pertanyaan dibangun berdasarkan tipe pertanyaannya. Dalam membangun pola pertanyaan ada beberapa tahapan yang harus diperhatikan (Toba, dkk, 2009) yaitu:

1. Kata tanya utama merupakan indikasi dari masing-masing tipe pertanyaan.
2. Posisi kata kunci dari pertanyaan dalam sebuah kalimat tanya menunjukkan konteks dan target dari pertanyaan.
3. Alternatif dari kata tanya atau frase, menunjukkan arti khusus untuk tipe pertanyaan tertentu.

Berikut contoh dalam membuat pola pertanyaan sesuai tahapan di atas untuk *property* tempat(Toba,dkk, 2009).

Tahap 1:

(dimana|dimanakah) letak <TO><CO>

(dimana|dimanakah) <CO> <TO>

(dimana|dimanakah) <TO> berada

Tahap 2:

<TO> ada dimana (saja)?

<TO> <PO> ada dimana

dimana (saja)? (kah|sih)? <TO> berada

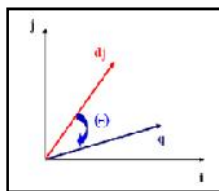
<TO> terletak <CO> (apa)?
 <TO> <CO> (ada)? dimana (saja)?
 <CO> <TO> (ada)? dimana (saja)?
 <TO> <CO> ada dimana?

Tahap 3:

(apa|apakah) nama <TO> <CO>
 (apa|apakah) nama (daerah|tempat|lokasi) <TO> <CO>
 (apa|apakah) nama (daerah|tempat|lokasi) <CO> <TO>
 di (lokasi|daerah|tempat) mana (saja)? <TO> <CO>
 di (lokasi|daerah|tempat) mana saja? <CO> <TO>
 terletak dimana (saja)? <TO> <CO>

2.5. Model Ruang Vektor

Model ruang vektor adalah salah satu pemodelan dalam sistem temu kembali informasi (IR *System*) yang sering digunakan. Model ini diperkenalkan oleh Salton untuk memperbaiki kekurangan dari model *boolean*. Model ini mampu menghasilkan dokumen-dokumen terurut berdasarkan tingkat relevansinya dengan *query* pengguna. Selain itu, *query* yang digunakan pengguna dapat dinyatakan dalam himpunan kata-kata atau istilah dengan bahasa sehari-hari. Di samping kelebihan yang telah dipaparkan di atas, model ini juga memiliki keterbatasan dimana *term-term* yang memiliki makna sama dapat dianggap sebagai elemen vektor yang berbeda. (Mandala, 2006). Pada pemodelan ini dokumen-dokumen direpresentasikan dengan vektor-vektor pada ruang multidimensi yang diilustrasikan pada Gambar 2.9.



Gambar 2.9. Ilustrasi *Similarity* 2 Dimensi

(Sumber: Gunawan, dkk, 2012)

Pada pemodelan ruang vektor, terdapat tiga tahapan dalam membangun IR *system* yaitu:

1. Pengindeksan dokumen
2. Pembobotan untuk menghasilkan dokumen yang relevan
3. Perangkingan berdasarkan ukuran kesamaan (*measure similarity*)

2.5.1. Pengindeksan Dokumen

Pengindeksan dokumen memiliki beberapa tahapan, yaitu:

a. Penguraian (*Parsing*)

Parsing adalah sebuah proses penguraian kalimat menjadi kata-kata atau frase-frase. *Parsing* akan menghasilkan daftar istilah dengan informasi tambahan seperti frekuensi dan posisi yang akan digunakan pada tahapan selanjutnya. Tahapan ini juga menghilangkan karakter-karakter tertentu dan semua kata akan diubah ke bentuk huruf kecil (*lower case*).

b. *Stopword Removing*

Stopword removing adalah proses menyaring kata dengan mengambil kata-kata penting dan menghapus kata-kata yang tidak penting yang terdapat dalam *stoplist*. Contoh kata yang termasuk ke dalam *stopwords* adalah pada, dari, di-, ke-, hanya dan sebagainya.

c. *Stemming*

Stemming adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya. Proses *stemming* bertujuan untuk meningkatkan kualitas informasi yang didapatkan. Untuk algoritma *stemming* pada bahasa Indonesia yang banyak digunakan adalah algoritma Porter dan algoritma Nazief & Adriani. Namun pada rancang bangun aplikasi QA System ini algoritma *stemming* yang digunakan adalah algoritma Nazief & Adriani. Hal ini dikarenakan algoritma tersebut memiliki nilai presisi yang lebih besar dibanding algoritma lainnya.

2.5.2. Pembobotan Kata (*Term Weighting*)

Dalam pembobotan kata terdapat dua hal penting yang menjadi pertimbangan yaitu *Term frequency* (tf) dan *Inverse Document Frequency* (idf)

Term frequency (tf) merupakan frekuensi kemunculan *term* (t) pada dokumen (d). Bobot dari *term* t pada dokumen d yaitu:

$$tf(d,t) = f(d, t) \dots\dots\dots(2.6)$$

Nilai jumlah kemunculan suatu kata (*term frequency*) diperhitungkan dalam pemberian bobot terhadap suatu kata. Semakin besar jumlah kemunculan suatu *term* (tf tinggi) dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar.

Inverse Document Frequency (idf) adalah pengurangan dominasi *term* yang sering muncul di berbagai dokumen. Pembobotan ini menganggap *term* yang jarang muncul pada kumpulan dokumen sangat bernilai. Kepentingan tiap *term* diasumsikan memiliki perbandingan yang berkebalikan dengan jumlah dokumen yang mengandung *term*. Faktor IDF dari *term* t dalam n buah dokumen yaitu :

$$idf(t) = \log(n / df(t)) \dots\dots\dots(2.7)$$

dimana:

df(t): jumlah dokumen yang mengandung *term* t.

n= banyak dokumen

Perkalian antara tf dan idf dapat menghasilkan performansi yang lebih baik. Kombinasi bobot dari *term* t pada dokumen d yaitu :

$$tfidf(d,t) = tf(d,t) \times idf(t) \dots\dots\dots(2.8)$$

Term yang sering muncul pada dokumen tertentu, tapi jarang muncul pada kumpulan dokumen, memberikan nilai bobot yang tinggi. Tfidf akan meningkat dengan jumlah kemunculan *term* pada dokumen tertentu, tapi akan berkurang dengan jumlah *term* yang muncul pada banyak dokumen.

2.5.3. Perangkingan berdasarkan ukuran kesamaan (*measure similarity*)

Salah satu ukuran kemiripan teks yang populer adalah *cosine similarity*. Ukuran ini menghitung nilai cosinus sudut antara dua vektor. Jika terdapat dua vektor dokumen d_j dan *query* Q , serta t *term* diekstrak dari koleksi dokumen maka nilai cosinus antara d_j dan Q didefinisikan sebagai:

$$\text{similarity} \left(\vec{d_j}, \vec{q} \right) = \frac{\vec{d_j} \cdot \vec{q}}{\|\vec{d_j}\| \cdot \|\vec{q}\|} = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{ij}^2 \cdot \sum_{i=1}^t w_{iq}^2}} \dots\dots\dots(2.9)$$

dimana:

- d_j : vektor dokumen j
- q : vektor *query*
- w_{ij} : bobot *term* i dalam dokumen j
- w_{iq} : bobot *term* i dalam *query*

2.6. Representasi Dokumen

Pengolahan dokumen yang digunakan pada rancang bangun QA *System* ini menggunakan bentuk standar yang terdiri dari *tag-tag* yang mengandung isi dari dokumen tersebut. Berdasarkan *corpus* (Adisantoso, 2004) dalam (Nurdiati, dkk, 2010) bentuk standar dokumen dalam koleksi adalah sebagai berikut:

```
<doc>
<doc no>...</doc no>
<title>...</title>
<author>...</author>
<docsource>...</docsource>
<date>...</date>
<abstract>...</abstract>
<keyword><kw>...</kw></keyword>
<text>
<caption>...</caption>
... (isi dokumen) ...
</text>
</doc>
```

Gambar 2.10. Bentuk Standar *Corpus*

(Sumber: Nurdiati, dkk, 2010)

2.7. Ekspresi Regular (*Regular Expression*)

Ekspresi regular merupakan sebuah ekspresi yang memberikan suatu pola (*pattern*) atau *template* untuk untaian/string dari suatu bahasa. Ekspresi regular memiliki beberapa kelebihan diantaranya:

- Sangat baik digunakan untuk mengelola dan mencocokkan *file* teks atau *string*.
- Sangat ringkas, karena sintaks Ekspresi Regular sangat “sederhana” untuk melakukan hal yang “besar”.
- Ekspresi Regular cepat, karena dapat menghindari pemeriksaan manual dan Ekspresi Regular cepat dalam mencocokkan pola-polanya.

2.7.1. Notasi Ekspresi Regular

Notasi Ekspresi Regular terdiri dari dua jenis karakter dasar, yaitu :

- Karakter teks literal (normal)

Karakter teks literal adalah karakter-karakter biasa yang tidak memiliki arti khusus atau karakter-karakter yang diterjemahkan apa adanya.

- Metakarakter

Metakarakter adalah karakter-karakter yang memiliki arti khusus dari Ekspresi Regular dan mewakili sekelompok karakter lain atau pola khusus tertentu. Berikut adalah tabel beberapa metakarakter pada Ekspresi Regular :

Tabel 2.3 Metakarakter pada Ekspresi Regular

Metakarakter	Keterangan
^....\$	Tanda awal dan akhir dari sebuah baris
{.....}	Untuk mendefinisikan <i>range</i>
[.....]	Untuk mendefinisikan <i>character class</i>
(.....)	Untuk mendefinisikan <i>group</i>
*	Untuk menyatakan perulangan minimal 0 kali sampai berhingga kali (0-n)
+	Untuk menyatakan perulangan minimal satu kali (1-n)
+ atau â	Berarti <i>union</i> (gabungan)
. (titik)	Berarti konkatenasi, biasanya tidak bias dihilangkan

Tabel 2.3 Metakarakter pada Ekspresi Regular (Lanjutan)

Metakarakter	Keterangan
?	Untuk menyatakan perulangan satu kali atau tidak sama sekali
	Operator logika untuk OR
\	Menjadikan karakter setelahnya tidak lagi dianggap sebagai metakarakter dan hanya akan dianggap sebagai karakter biasa
\t	Mereperesentasikan karakter <i>tab</i>
\n	Mereperesentasikan karakter <i>newline</i>
\r	Mereperesentasikan <i>carriage-return character</i>
\f	Mereperesentasikan <i>form-feed character</i>
\a	Mereperesentasikan karakter bel
\e	Mereperesentasikan karakter <i>escape</i>

Selain itu, juga terdapat kelas karakter yang dapat digunakan untuk mencari kesesuaian atau kecocokan karakter pada pola yang telah ditentukan.

Tabel 2.4 Kelas Karakter Pada Ekspresi Regular

Kelas Karakter	Keterangan
[abc]	Hanya a atau b atau c
[^abc]	Semua karater selain a , b dan c
[a-zA-Z]	Salah satu karakter dari a sampai z atau dari A sampai Z
[a-d[m-p]]	Salah satu karakter dari hasil gabungan antara a samapi d dan m sampai p, (union = [a-dm-p])
[a-z&&[def]]	Salah satu karakter dari hasil irisan a sampai z dan def (intersection = [def])
[a-z&&[^bc]]	Salah satu karakter dari a sampai z selain b dan c (subtraction = [ad-z])

Kelas karakter pada Ekspresi Regular yang telah didefinisikan diantaranya adalah :

Tabel 2.5 Kelas Karakter Pada Ekspresi Regular yang telah didefinisikan

Kelas Karakter	Keterangan
.	Mewakili semua karakter kecuali line terminator
\d	Merepresentasikan sebuah angka = [0-9]
\D	Merepresentasikan selain angka = [^0-9]
\s	Merepresentasikan karakter whitespace = [\t\n\f\r\x0B]
\S	Merepresentasikan selain karakter whitespace = [^\s]
\w	Merepresentasikan sebuah huruf atau angka = [a-zA-Z0-9_]
\W	Merepresentasikan selain huruf atau angka = [^\w]

2.8. Tata Bahasa Indonesia

Seperti kebanyakan bahasa lainnya, bahasa Indonesia juga memiliki tata bahasa. Berikut akan dijelaskan mengenai tata bahasa Indonesia yang digunakan dalam penelitian ini.

2.8.1. Kalimat Tanya (Interogatif)

Kalimat tanya adalah kalimat yang mengharapkan adanya jawaban secara verbal. Jawaban tersebut dapat berupa keterangan, alasan, pengakuan, atau pendapat dari pihak pendengaran atau pembaca. (Chaer, 2009).

Dalam kalimat tanya terdapat kata tanya yang menunjukkan jenis dari pertanyaan tersebut. Kata tanya tersebut antara lain adalah apa/apakah, dimana/dimanakah, bagaimana/bagaimanakah, mengapa/ mengapa, kapan/kapankah, dan siapa/siapa. Berikut dijelaskan mengenai jenis dan pemakaian kalimat tanya.

- a. Kalimat tanya yang meminta pengakuan jawaban “ya” atau “tidak”, kalimat tersebut dapat dibentuk dengan cara member intonasi tanya pada sebuah klausa (kalimat).
- b. Kalimat tanya yang meminta jawaban mengenai salah satu unsur kalimat dibentuk dengan bantuan kata tanya apa, siapa, mana, berapa, dan kapan sesuai dengan bagian mana dari kalimat yang ingin ditanyakan.
 - 1) Apa, untuk menanyakan benda, pengertian, tujuan, manfaat.
 - 2) Siapa, untuk menanyakan orang atau makhluk hidup lain yang melakukan sesuatu.
 - 3) Mana, untuk menanyakan keberadaan benda (termasuk orang). Untuk menanyakan tempat keberadaan, tempat kedatangan, dan tempat tujuan dengan lebih tepat di depan kata mana ditambahkan preposisi di, ke, dan dari.
 - 4) Berapa, untuk menanyakan jumlah atau banyaknya sesuatu yang digunakan.

- 5) Kapan, untuk menanyakan waktu terjadi suatu peristiwa. Untuk menanyakan permulaan terjadinya sesuatu harus digunakan kata tanya sejak kapan, dan untuk menanyakan batas akhir akan terjadinya sesuatu harus digunakan kata tamua sampai kapan.
- 6) Mengapa, untuk menanyakan tujuan

2.9. Evaluasi *Question Answering (QA) System*

QA *System* bertujuan untuk menemukembalikan informasi yang relevan dan memperkecil kemungkinan untuk menemukembalikan informasi yang tidak relevan. Dalam menilai relevan atau tidaknya informasi yang ditemu kembalikan oleh sistem dapat diukur dari benar atau tidaknya jawaban yang dikembalikan oleh sistem kepada pengguna. Untuk mengetahui benar atau tidaknya jawaban tersebut, diberlakukan beberapa ketentuan. Ketentuan-ketentuan tersebut (Dang, dkk, 2007) yaitu:

- 1) *incorrect* : jawaban tidak benar atau jawabannya tidak responsif.
- 2) *not supported* : jawaban benar namun dokumen yang dikembalikan tidak mendukung jawaban tersebut.
- 3) *not exact* : jawaban dan dokumen yang dikembalikan benar, namun terlalu panjang.
- 4) *correct* : jawaban dan dokumen yang dikembalikan benar.

Jawaban-jawaban yang dikembalikan sebagai keluaran dari pertanyaan yang dimasukkan oleh pengguna tersebut dijadikan evaluasi untuk melihat performansi dari QA *System* yang telah dibangun. Metode presisi (*precision*) dan *confidence weighted score* (CWS) digunakan sebagai parameter untuk mengukur performansi dari sistem. Berikut penjelasan lebih lanjut.

a. Presisi (*Precision*)

Presisi yang dihitung adalah presisi berdasarkan jawaban yang benar (*precision of correct*) dan presisi berdasarkan jawaban tidak benar (*precision of incorrect*).

$$Precision\ of\ Correct = \frac{jumlah\ jawaban\ yang\ benar}{jumlah\ jawaban\ yang\ ditemukan} \times 100\% \dots\dots\dots(2.10)$$

$$Precision\ of\ Incorrect = \frac{jumlah\ jawaban\ yang\ salah}{jumlah\ jawaban\ yang\ ditemukan} \times 100\% \dots\dots\dots(2.11)$$

b. *Confidence Weighted Score*

Confidence Weighted Score dapat dirumuskan sebagai berikut:

$$CWS = \frac{1}{Q} \sum_{i=1}^Q \frac{\#correct\ in\ first\ i}{i} \dots\dots\dots(2.12)$$

dimana:

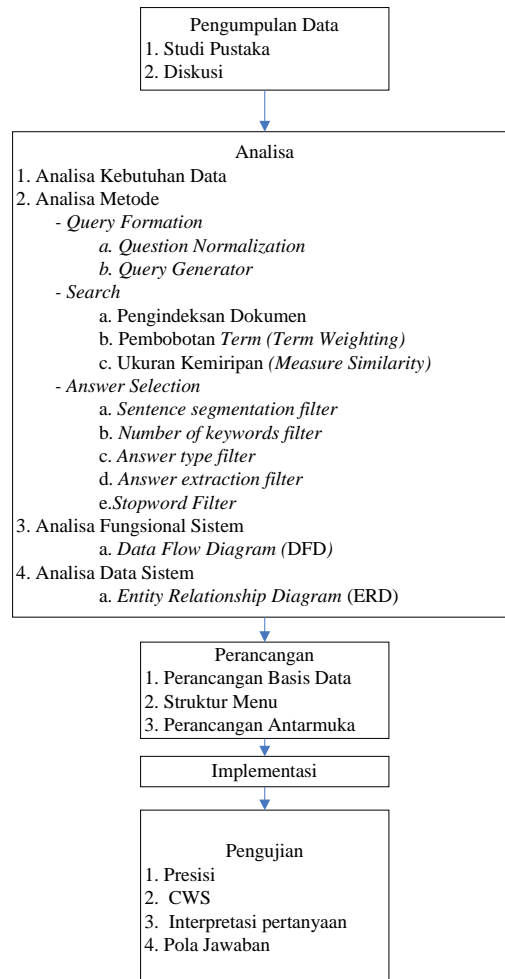
- | | |
|---------------------------|--|
| Q | : Jumlah pertanyaan |
| i | : Jumlah jawaban yang dikembalikan |
| <i>correct in first i</i> | : Urutan jawaban yang benar pertama kali ditemukan |

BAB III

METODOLOGI PENELITIAN

3.1. Alur Metodologi Penelitian

Dalam penelitian Tugas Akhir ini ada lima tahapan yang akan penulis lakukan, yaitu: pengumpulan data, analisa, perancangan, implementasi dan pengujian aplikasi. Alur metodologi penelitian tersebut dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Alur Metodologi Penelitian

3.1.1. Pengumpulan Data

Pada tahap pengumpulan data, penulis menggunakan dua pendekatan guna memperoleh informasi-informasi untuk dapat merancang bangun QA *System* tersebut. Pendekatan tersebut dapat diuraikan sebagai berikut:

1. Studi Pustaka

Studi pustaka yang penulis gunakan pada penelitian ini adalah mengumpulkan teori-teori terkait rancang bangun QA *System* ini tentang QA *System* dan penerapannya, Ephyra *Framework*, tata bahasa Indonesia, dan terjemahan Al Quran yang diambil dari buku-buku referensi, Tugas Akhir, maupun jurnal-jurnal nasional dan internasional.

2. Diskusi

Merupakan metode yang dimaksudkan untuk berdiskusi dalam menyelesaikan permasalahan yang dibahas dalam laporan ini dengan orang-orang yang memahami tentang kasus pembahasan atau berdiskusi tentang masalah perancangan aplikasi yang akan dibangun.

3.1.2. Analisa Sistem

Analisa sistem dilakukan setelah dilakukannya pengumpulan data. Analisa dilakukan untuk mengidentifikasi seluruh masalah yang ada pada penelitian. Dalam penelitian ini analisa yang dilakukan meliputi beberapa tahapan yaitu:

- a. Menganalisa data-data yang dibutuhkan oleh sistem.
- b. Analisa Metode, yang meliputi:
 1. Pembentukan *query* (*query formation*) yang digunakan untuk membentuk *query* dari kalimat pertanyaan dari masukan pengguna. Tahapan ini terdiri dari:
 - a) *Question normalization*, yang berfungsi untuk normalisasi pertanyaan.
 - b) *Query generator* yang digunakan untuk menghasilkan *query* yang akan digunakan pada tahap pencarian.

2. Pencarian (*search*), digunakan untuk menemukan koleksi dokumen yang relevan dengan *query*. Tahapan ini terdiri dari: pengindeks-an dokumen yang meliputi *parsing*, *stopword removal*, dan *stemming*; pembobotan kata, dan mengukur kemiripan.
3. Seleksi Jawaban (*answer selection*) digunakan untuk menyeleksi dari dokumen dan menampilkan jawaban yang sesuai kepada pengguna.
 - a) Tahapan ini terdiri dari segmentasi kalimat, menghitung jumlah kata kunci, penyaringan berdasarkan tipe jawaban, ekstraksi jawaban dan *stopword filter*.
 - b) Analisa Fungsional Sistem, yaitu menganalisis fungsional aplikasi dengan menggunakan *Data Flow Diagram* (DFD).
 - c) Analisa data sistem, yaitu menganalisis data yang diperlukan pada aplikasi dengan menggunakan *Entity Relationship Diagram* (ERD).

3.1.3. Perancangan Sistem

Setelah selesai melakukan tahap analisa selajutnya dilakukan perancangan berdasarkan hasil analisa tersebut. Perancangan ini mencakup beberapa tahapan yaitu:

- a. Perancangan basis data, yaitu perancangan basis data yang dibutuhkan dalam membangun sistem.
- b. Perancangan Struktur Menu
- c. Perancangan antarmuka sistem, yaitu bagaimana membuat tampilan yang *user friendly* bagi pengguna.

3.1.4. Implementasi

Implementasi merupakan metode yang digunakan setelah analisa dan perancangan rancang bangun aplikasi selesai dilakukan. Metode ini akan menjelaskan tentang penerapan jalannya rancang bangun yang telah dianalisa dan dirancang. Implementasi pengembangan aplikasi ini akan dikembangkan pada spesifikasi *hardware* dan *software* berikut:

1. Perangkat keras

Processor : *Intel(R) Core(TM) i5-2450M CPU @ 2.50GHz*

Memory : 8.00 GB

2. Perangkat Lunak

Sistem operasi : *Windows 7 Home Premium 64-bit (6.1,Build 7601)*

Bahasa pemrograman : *PHP Script Language Version 5.2.6.*

Tools perancangan : Notepad ++ versi 5.9.3.

Browser : Google Chrome, Mozilla FireFox

DBMS : *mySQL*

Web server : *Apache Web Server Version 2.2.8.*

3.1.5. Pengujian

Adapun tahapan pengujian yang akan dilakukan pada aplikasi *QA System* yang telah dibangun meliputi:

1. Perhitungan nilai presisi.
2. Perhitungan CWS
3. Interpretasi pertanyaan
4. Pola jawaban

3.1.6. Kesimpulan dan Saran

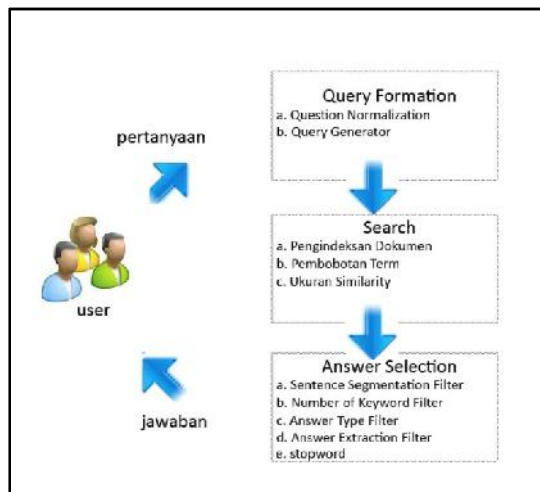
Pada bagian ini, berisi kesimpulan mengenai hasil evaluasi dari seluruh kegiatan yang dilakukan dalam melakukan penelitian terhadap Analisa dan Penerapan *Ephyra framework* dalam aplikasi *QA System* terjemahan Al Quran. Selain itu juga dijelaskan mengenai performansi dari sistem. Pada tahap ini juga diberikan saran-saran untuk pengembangan dan pengelolaan sistem lebih lanjut.

BAB IV

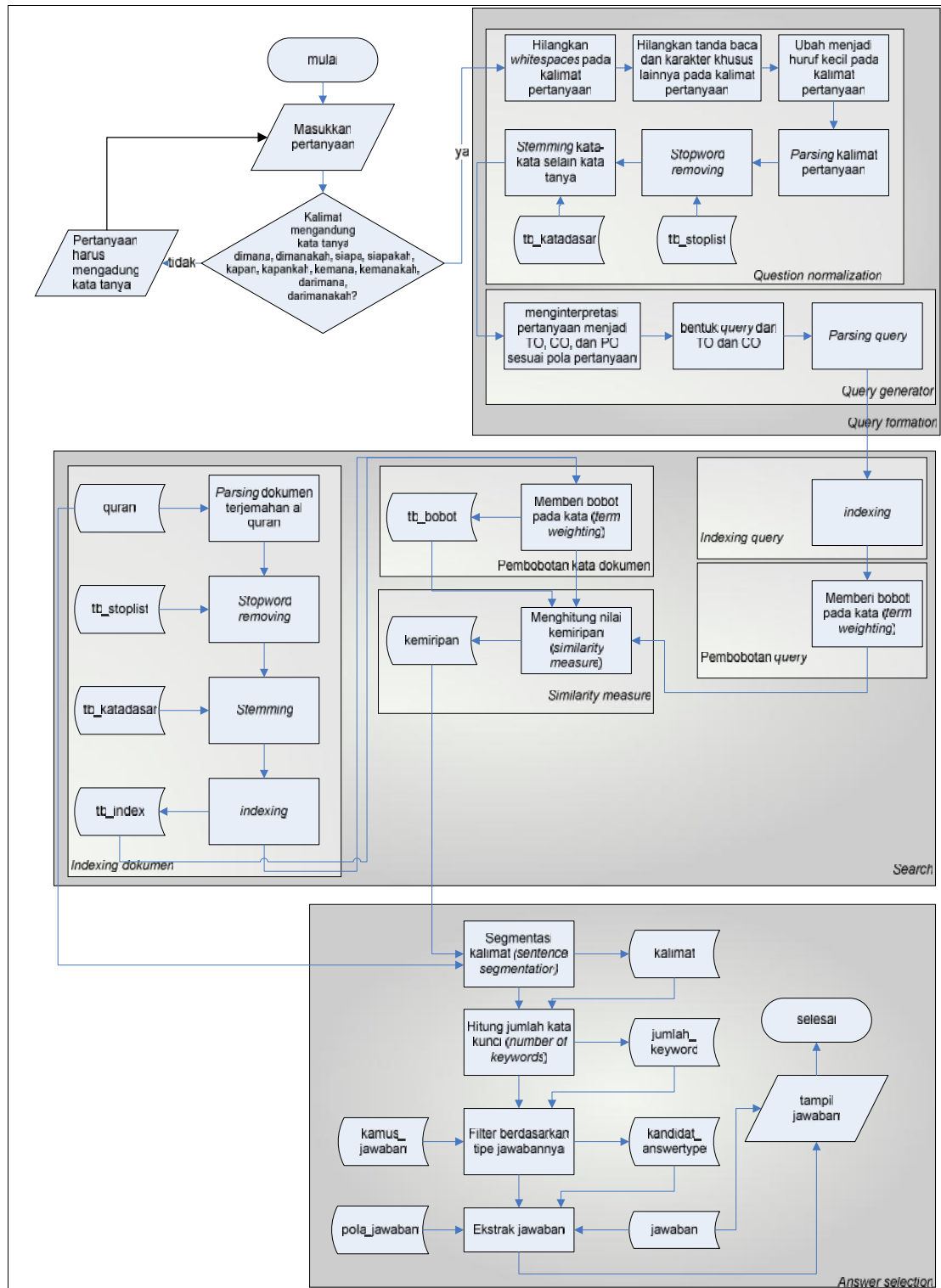
ANALISA DAN PERANCANGAN

4.1. Deskripsi Umum Sistem

Aplikasi QA System yang dibuat secara umum memiliki 3 tahapan yang diadaptasi dari arsitektur Ephyra yaitu, pembentukan *query* (*query formation*), pencarian (*search*), dan seleksi jawaban (*answer selection*). Pada pembentukan *query* (*query formation*) terdapat 2 tahapan utama yaitu normalisasi pertanyaan (*question normalization*) dan pembentukan *query* (*query generator*); sedangkan pada pencarian (*search*) menggunakan pemodelan ruang vektor dimana terdapat 3 tahapan utama yaitu pengindeksan (*indexing*), pembobotan kata (*term weighting*), serta pengukuran kemiripan (*similarity measure*); dan untuk seleksi jawaban (*answer selection*) terdapat 5 tahapan utama yaitu *sentence segmentation filter*, *number of keywords filter*, *answer type filter*, *answer extraction filter* dan *stopword filter*. Untuk lebih jelasnya Gambar 4.1 menggambarkan sistem secara keseluruhan.



Gambar 4.1. Deskripsi Umum Sistem



Gambar 4.2. Flowchart Proses secara Keseluruhan

Pada mulanya sistem menerima *input* dari *user* berupa kalimat pertanyaan. Sistem kemudian menganalisis dari pertanyaan yang di-*inputkan* *user* dan mengubahnya menjadi *query*. Setelah itu, *query* tersebut akan digunakan pada tahapan pencarian. Sistem melakukan pencarian menggunakan model ruang vektor dan mengurutkan dokumen-dokumen yang ditemu kembalikan berdasarkan nilai kemiripannya. Setelah itu sistem akan menyeleksi kandidat-kandidat jawaban yang ditemukan dan menampilkan kandidat jawaban terbaik ditampilkan kepada *user*. Gambar 4.2 di atas menunjukkan *flowchart* sistem secara keseluruhan.

4.1.1 Analisa Kebutuhan Data

Pada pembuatan aplikasi QA System ini data yang dibutuhkan dapat dijabarkan secara umum sebagai berikut:

a. Pengguna

Pengguna berisi data *administrator* yang berperan sebagai pengelola sistem. Data ini digunakan untuk proses verifikasi agar dapat mengakses sistem. Data ini mencakup *username*, dan *password*.

b. Kata tanya

Kata tanya berisi data kata tanya-kata tanya yang akan digunakan. Data ini mencakup kata tanya.

c. *Stoplist*

Stoplist berisi data kata-kata yang tidak begitu berpengaruh atau tidak memiliki arti yang penting pada *query* pertanyaan, dokumen maupun jawaban. Data ini mencakup *stoplist*.

d. Kata dasar dalam bahasa Indonesia

Kata dasar berisi data mengenai bentuk kata dasar dan tipe kata tersebut. Data ini mencakup kata dasar dan tipe kata dasar.

e. Al Quran

Al Quran berisi data mengenai informasi terjemahan Al Quran yang mencakup id dokumen, teks bahasa arab, dan nama *file* terjemahan.

f. Indeks

Indeks berisi data hasil pengindeksan dokumen yang mencakup id, *term*, id dokumen, dan frekuensi.

g. Bobot

Bobot berisi data hasil pembobotan dokumen yang mencakup *term*, id dokumen, dan nilai bobot dari *term*.

h. Kemiripan

Kemiripan berisi data mengenai dokumen-dokumen yang memiliki kemiripan dengan *query* yang mencakup id dokumen yang memiliki kemiripan dengan *query*, dan nilai kemiripannya.

i. Kalimat

Kalimat berisi data mengenai kalimat yang telah disegmentasi pada tahap segmentasi kalimat yang mencakup id kalimat, id dokumen, surat, ayat, kalimat dan hasil kalimat yang telah dilakukan *parsing*, *stopword removing*, dan *stemming*.

j. Jumlah *keyword*

Jumlah *keyword* berisi jumlah *keyword* dari kalimat yang menjadi kandidat jawaban. Jumlah *keyword* mencakup id kalimat dan jumlah *keyword*.

k. Kandidat *answer type*

Kandidat *answer type* berisi data kalimat yang menjadi kandidat jawaban yang dihasilkan pada tahap *answer type filter*. Data ini mencakup id kalimat dan nilai estimasi dari kalimat.

l. Kamus Jawaban

Kamus jawaban berisi data mengenai jawaban yang terdapat dalam terjemahan Al Quran yang mencakup jawaban, dan tipe jawaban.

m. Pola Jawaban

Pola jawaban berisi data mengenai pola jawaban yang digunakan untuk mengekstrak *property*. Data ini mencakup id pola, tipe, pola jawaban, jumlah benar dan jumlah salah.

n. Jawaban

Jawaban berisi data yang menjadi kandidat jawaban. Data ini mencakup id kalimat, id pola, jawaban, dan nilai *confidence* dari jawaban.

4.1.2 Analisa Metode

Seperti yang telah dijelaskan pada bab-bab sebelumnya mengenai Ephyra *framework*, pada subbab ini akan diuraikan mengenai alur kerja dari aplikasi QA *System* yang akan dibangun sesuai dengan kerangka kerja dari Ephyra dengan pola bahasa Indonesia pada pemodelan ruang vektor. Pada subbab analisa ini untuk mempermudah pemahaman mengenai proses-proses yang terdapat pada pembangunan QA *System* digunakan sebuah contoh pertanyaan yaitu:

Pertanyaan: “*Siapakah pencipta langit dan bumi?*”

dengan tiga buah korpus yaitu:

Dokumen 1 (dok1) : Segala puji bagi Allah, Tuhan semesta alam.

Dokumen 2(dok2) : Allah pencipta langit dan bumi dan bila Dia berkehendak untuk menciptakan sesuatu, maka (cukuplah) Dia hanya mengatakan kepadanya: “*Jadilah!*” lalu jadilah ia.

Dokumen 3 (dok3) : Sesungguhnya dalam penciptaan langit dan bumi, silih bergantinya malam dan siang, bahtera yang berlayar di laut membawa apa yang berguna bagi manusia, dan apa yang Allah turunkan dari langit berupa air, lalu dengan air itu dia hidupkan bumi sesudah mati (kering)-Nya dan dia sebarkan di bumi itu segala jenis hewan, dan pengisaran angin dan awan yang dikendalikan antara langit dan bumi; sungguh (terdapat) tanda-tanda (keesaan dan kebesaran Allah) bagi kaum yang memikirkan.

4.1.2.1 Pembentukan *Query* (*Query Formation*)

Pembentukan *query* bertujuan untuk membentuk *query* dari pertanyaan yang dimasukkan oleh *user* agar dapat diproses lebih lanjut. Pembentukan *query* memiliki dua tahapan utama yaitu normalisasi pertanyaan (*question normalization*), dan generator *query* (*query generator*). Masing-masing tahapan pada alur kerja yang digunakan dalam pembangunan QA System ini dijelaskan sebagai berikut:

a. *Question Normalization*

Question Normalization berfungsi untuk melakukan normalisasi terhadap kalimat pertanyaan (*question string*) yang dimasukkan. Proses yang dilakukan adalah:

1. Pada tahap awal dilakukan penghilangan *whitespaces* seperti *break line* dan lain-lain. Maka pertanyaan di atas akan menjadi:
Siapakah pencipta langit dan bumi?
2. Selanjutnya menghilangkan tanda baca dan karakter khusus lainnya.
Siapakah pencipta langit dan bumi
3. Setelah karakter-karakter tersebut dihilangkan, kemudian huruf-huruf pada kalimat pertanyaan diubah menjadi huruf kecil (*lower case*).
siapakah pencipta langit dan bumi
4. Tahap selanjutnya melakukan penguraian (*parsing*) pada kalimat pertanyaan.
siapakah
pencipta
langit
dan
bumi
5. Menghilangkan token-token yang tidak perlu (*stopword removing*). *Stopwords* yang digunakan disimpan pada tabel *tb_stoplist*. Daftar *stopwords* dapat dilihat pada Lampiran A. Hasil dari pertanyaan di atas adalah:
siapakah pencipta langit bumi

6. Menjadikan verba dan nomina menjadi kata dasar (*stemming*). *Stemming* yang digunakan pada rancang bangun QA System ini adalah algoritma Nazief dan Adriani. Hasil dari proses ini untuk pertanyaan di atas adalah:

siapakah cipta langit bumi

b. *Query Generator*

Pada tahap ini kalimat pertanyaan yang telah dinormalisasi akan diubah menjadi sebuah *query* yang diteruskan ke tahap selanjutnya. Tahapan pada proses ini sebagai berikut:

1. Interpretasi Pertanyaan

Seperti yang dijelaskan pada Bab II, interpretasi pertanyaan adalah tahapan untuk menginterpretasikan pertanyaan menjadi tiga buah objek yaitu: *property* (PO), target (TO), dan *context* (CO). PO, TO, dan CO dapat dihasilkan dari pencocokan pola-pola pertanyaan (*question patterns*) yang telah didefinisikan secara manual terlebih dahulu saat membangun sistem. Pola pertanyaan (*question patterns*) berdasarkan kategori orang, waktu, dan tempat yang dibangun pada aplikasi QA System ini dapat dilihat pada Tabel 4.1 di bawah ini.

Tabel 4.1 Pola Pertanyaan (*Question Patterns*) untuk *Property* Orang, Tempat, Waktu

No	<i>Property</i>	Pola Pertanyaan (<i>Question Patterns</i>)
1	Orang	(siapa siapakah) <TO> <CO>
2	Orang	<TO> (siapa siapakah) <CO>
3	Orang	<CO> <TO> (siapa siapakah)
4	Tempat	(dimana dimanakah) (letak tempat) <TO> <CO>
5	Tempat	(dimana dimanakah) <CO> <TO>
6	Tempat	(kemana kemanakah) <CO> <TO>
7	Tempat	(darimana darimanakah) <TO> <CO>
8	Tempat	(dimana dimanakah) <TO> ada
9	Waktu	(kapan kapankah) <TO> <CO>

Untuk contoh pertanyaan “Siapakah pencipta langit dan bumi?” di atas, maka pola pertanyaan yang sesuai adalah pola (siapa|siapakah) <TO> <CO> sehingga didapatkan:

Target (TO): cipta

Context (CO): langit bumi

Property (PO): orang

2. *Query*

Setelah PO, TO, dan CO ditentukan maka TO dan CO tersebut akan menjadi *query* untuk diproses ke tahap selanjutnya. *Query* : cipta langit bumi.

Setelah *query* didapat, maka *query* tersebut diuraikan (*parsing*) agar dapat diproses pada tahap pencarian. Hasil penguraian (*parsing*) dari *query* di atas adalah sebagai berikut:

cipt

langit

bumi

4.1.2.2 Pencarian (*Search*)

a. Koleksi Dokumen

Dokumen yang digunakan pada penelitian ini berformat teks (.txt). Nama dokumen berdasarkan urutan surat dan ayat dari Al Quran dengan jumlah digit angka yang digunakan pada nama dokumen adalah empat, sehingga pemberian nama dokumen pada surat Al Fatihah ayat 1 adalah 0001 dan seterusnya. Dokumen-dokumen tersebut mengandung *tag-tag* XML dan disimpan dalam sebuah direktori. Dokumen-dokumen tersebut terdiri dari *tag-tag*: <DOC></DOC>, <DOCNO></DOCNO>, <SURAT></SURAT>, <AYAT></AYAT>, <TERJEMAHAN></TERJEMAHAN>.

Tiap-tiap dokumen berisi satu terjemahan ayat yang diakhiri dengan tanda titik (.) di setiap akhir kalimat. Berikut merupakan contoh format dokumen yang digunakan (Lihat Gambar 4.3).

```

<DOC>
<DOCNO>
    0001
</DOCNO>
<SURAT>
    Al-Fatihah
</SURAT>
<AYAT>
    1
</AYAT>
<TERJEMAHAN>
    Dengan menyebut nama Allah Yang Maha Pemurah lagi Maha
    Penyayang.
</TERJEMAHAN>
</DOC>

```

Gambar 4.3. Format Koleksi Dokumen Sistem

Pada beberapa dokumen yang digunakan dalam penelitian ini, terdapat penggabungan dari beberapa ayat. Hal ini dikarenakan ayat-ayat tersebut diakhiri dengan tanda koma (,) dan ayat tersebut merupakan lanjutan dari ayat sebelumnya. Penggabungan tersebut dilakukan secara manual oleh penulis. Daftar dokumen yang mengalami penggabungan ayat dapat dilihat pada Lampiran B. Selain itu, ayat-ayat yang diakhiri tanda tanya (?) dan tanda seru (!) yang tidak berhubungan dengan ayat selanjutnya, penulis menambahkan tanda titik (.) secara manual diakhir kalimatnya yang dapat dilihat pada Lampiran C. Nama dokumen yang digunakan disimpan dalam tabel quran.

b. Pengindeksan (*Indexing*) Dokumen

Tahapan pengindeksan memiliki beberapa proses, yaitu:

1. Pada tahap awal dilakukan *parsing* (penguraian), yaitu dengan melakukan pemecahan kata pada dokumen. Hasil *parsing* untuk dokumen-dokumen di atas dapat dilihat pada Tabel 4.2, Tabel 4.3, dan Tabel 4.4.

Dokumen 1 (dok1):

Tabel 4.2 Hasil *Parsing* pada Dokumen 1

Segala
Puji
Bagi
Allah
Tuhan
Semesta
Alam

Dokumen 2 (dok2):

Tabel 4.3 Hasil *Parsing* pada Dokumen 2

allah	berkehendak	hanya
Pencipta	untuk	mengatakan
Langit	menciptakan	kepadanya
Dan	sesuatu	jadilah
Bumi	maka	lalu
Dan	cukuplah	jadilah
Dia	Dia	ia

Dokumen 3 (dok3):

Tabel 4.4 Hasil *Parsing* pada Dokumen 3

sesungguhnya	allah	dan
Dalam	turunkan	pengisaran
Penciptaan	dari	angin
Langit	langit	dan
Dan	berupa	awan
Bumi	Air	yang
Silih	Lalu	dikendalikan
Bergantinya	dengan	antara
Malam	Air	langit
Dan	Itu	dan
Siang	Dia	bumi
Bahtera	hidupkan	sungguh
Yang	bumi	terdapat
Berlayar	sesudah	tanda
Di	mati	tanda
Laut	keringnya	keesaan
Membawa	Dan	dan

Tabel 4.4 Hasil *Parsing* pada Dokumen 3 (Lanjutan)

Apa	Dia	kebesaran
Yang	sebarkan	allah
Berguna	Di	bagi
Bagi	bumi	kaum
Manusia	Itu	yang
Dan	segala	memikirkan
Apa	jenis	
Yang	hewan	

- Selanjutnya, kata-kata yang tidak begitu berpengaruh pada *query* pertanyaan akan dieliminasi dan kata-kata yang termasuk dalam *stopwords* atau sering dikenal dengan proses *stopword removing*. *Stopwords* tersebut disimpan pada tabel *tb_stoplist*. Daftar *stopwords* dapat dilihat pada Lampiran A dan hasil *stopword removing* untuk dokumen-dokumen di atas dapat dilihat pada Tabel 4.5, Tabel 4.6, dan Tabel 4.7.

Dokumen 1 (dok1):

Tabel 4.5 Hasil *Stopword Removing* pada Dokumen 1

Puji
Allah
Tuhan
Semesta
Alam

Dokumen 2 (dok2):

Tabel 4.6 Hasil *Stopword Removing* pada Dokumen 2

Allah	cukuplah	jadilah
Pencipta	hanya	
Langit	mengatakan	
Bumi	kepadanya	
berkehendak	jadilah	
menciptakan	Lalu	

Dokumen 3 (dok3):

Tabel 4.7 Hasil *Stopword Removing* pada Dokumen 3

Dalam	allah	hewan
Penciptaan	turunkan	pengisaran
Langit	langit	angin
Bumi	Air	awan
Silih	Lalu	dikendalikan
Bergantinya	Air	langit
Malam	Itu	bumi
Siang	hidupkan	sungguh
Bahtera	bumi	terdapat
Berlayar	sesudah	tanda
Di	mati	tanda
Laut	keringnya	keesaan
Membawa	sebarakan	kebesaran
Apa	Di	allah
Berguna	bumi	kaum
Manusia	Itu	memikirkan
Apa	jenis	

- Setelah itu, untuk kata-kata yang tersisa akan dilakukan proses *stemming*. Proses *stemming* pada penelitian ini menggunakan algoritma Nazief dan Adriani, untuk melakukan proses *stemming* kata dasar disimpan pada tabel tb_katadasar. Hasil *stemming* untuk dokumen-dokumen di atas dapat dilihat pada Tabel 4.8, Tabel 4.9, dan Tabel 4.10.

Dokumen 1 (dok1):

Tabel 4.8 Hasil *Stemming* pada Dokumen 1

Puji
Allah
Tuhan
Semesta
Alam

Dokumen 2 (dok2):

Tabel 4.9 Hasil *Stemming* pada Dokumen 2

allah	cukup	jadi
Cipta	hanya	
Langit	atak	
Bumi	kepadanya:	
berkehendak	jadi	
Cipta	Lalu	

Dokumen 3 (dok3):

Tabel 4.10 Hasil *Stemming* pada Dokumen 3

Dalam	Apa	itu	kaum
Cipta	Allah	jenis	memikirkan
Langit	Turun	hewan	
Bumi	Langit	pengisaran	
Silih	Air	angin	
Ganti	Lalu	awan	
Malam	Air	kendali	
Siang	Itu	langit	
Bahtera	Hidup	bumi	
Layar	Bumi	sebenarnya	
Di	Sudah	dapat	
Laut	Mati	tanda	
Bawa	Ring	tanda	
Apa	Bar	esa	
Guna	Di	besaran	
Manusia	Bumi	allah	

c. Pembobotan Kata (*Term Weighting*)

Dalam pembobotan kata (*term weighting*) menggunakan persamaan 2.6-2.8 dan disimpan pada tabel *tb_index* dan *tb_bobot*. Tabel 4.11 menjelaskan hasil perhitungan pembobotan kata untuk contoh di atas.

Tabel 4.11 Hasil Perhitungan Tf Idf pada Pembobotan Kata

<i>Term</i>	tf				df	d/df	idf= log(d/df)	W=tfxidf			
	Q	Dok 1	Dok 2	Dok 3				Q	Dok1	Dok2	Dok3
puji	0	1	0	0	1	3	0,477	0	0,477	0	0
allah	0	1	1	2	3	1	0	0	0	0	0
tuhan	0	1	0	0	1	3	0,477	0	0,477	0	0
semesta	0	1	0	0	1	3	0,477	0	0,477	0	0
alam	0	1	0	0	1	3	0,477	0	0,477	0	0
cipta	1	0	2	1	2	1,5	0,176	0,176	0	0,352	0,176
langit	1	0	1	3	2	1,5	0,176	0,176	0	0,176	0,528
bumi	1	0	1	4	2	1,5	0,176	0,176	0	0,176	0,704
berkehen- dak	0	0	1	0	1	3	0,477	0	0	0,477	0
cukup	0	0	1	0	1	3	0,477	0	0	0,477	0
hanya	0	0	1	0	1	3	0,477	0	0	0,477	0
atak	0	0	1	0	1	3	0,477	0	0	0,477	0
kepadanya:	0	0	1	0	1	3	0,477	0	0	0,477	0
jadi	0	0	2	0	1	3	0,477	0	0	0,954	0
lalu	0	0	1	1	2	1,5	0,176	0	0	0,176	0,176
dalam	0	0	0	1	1	3	0,477	0	0	0	0,477
silih	0	0	0	1	1	3	0,477	0	0	0	0,477
ganti	0	0	0	1	1	3	0,477	0	0	0	0,477
malam	0	0	0	1	1	3	0,477	0	0	0	0,477
siang	0	0	0	1	1	3	0,477	0	0	0	0,477
bahtera	0	0	0	1	1	3	0,477	0	0	0	0,477
layar	0	0	0	1	1	3	0,477	0	0	0	0,477
di	0	0	0	2	1	3	0,477	0	0	0	0,954
laut	0	0	0	1	1	3	0,477	0	0	0	0,477
bawa	0	0	0	1	1	3	0,477	0	0	0	0,477

Tabel 4.11 Hasil Perhitungan Tf Idf pada Pembobotan Kata (Lanjutan)

Term	tf				df	d/df	idf= log(d /df)	W=tfxidf			
	Q	Dok 1	Dok 2	Dok 3				Q	Dok 1	Dok 2	Dok 3
apa	0	0	0	2	1	3	0,477	0	0	0	0,954
guna	0	0	0	1	1	3	0,477	0	0	0	0,477
manusia	0	0	0	1	1	3	0,477	0	0	0	0,477
turun	0	0	0	1	1	3	0,477	0	0	0	0,477
air	0	0	0	1	1	3	0,477	0	0	0	0,477
itu	0	0	0	2	1	3	0,477	0	0	0	0,954
hidup	0	0	0	1	1	3	0,477	0	0	0	0,477
sudah	0	0	0	1	1	3	0,477	0	0	0	0,477
mati	0	0	0	1	1	3	0,477	0	0	0	0,477
ring	0	0	0	1	1	3	0,477	0	0	0	0,477
bar	0	0	0	1	1	3	0,477	0	0	0	0,477
jenis	0	0	0	1	1	3	0,477	0	0	0	0,477
hewan	0	0	0	1	1	3	0,477	0	0	0	0,477
pengisa- ran	0	0	0	1	1	3	0,477	0	0	0	0,477
angin	0	0	0	1	1	3	0,477	0	0	0	0,477
awan	0	0	0	1	1	3	0,477	0	0	0	0,477
kendali	0	0	0	1	1	3	0,477	0	0	0	0,477
sungguh	0	0	0	1	1	3	0,477	0	0	0	0,477
dapat	0	0	0	1	1	3	0,477	0	0	0	0,477
tanda	0	0	0	2	1	3	0,477	0	0	0	0,954
esa	0	0	0	1	1	3	0,477	0	0	0	0,477
besaran	0	0	0	1	1	3	0,477	0	0	0	0,477
kaum	0	0	0	1	1	3	0,477	0	0	0	0,477
memikir- kan	0	0	0	1	1	3	0,477	0	0	0	0,477
Total bobot									0	0,704	1,409

d. Ukuran Kemiripan (*Similarity Measure*)

Ukuran kemiripan (*similarity measure*) dihitung dengan menggunakan persamaan 2.9. Hasil perhitungan akan disimpan pada tabel kemiripan. Berdasarkan Tabel 4.11 dokumen yang mengandung *query* "cipta langit bumi" adalah dokumen 2

(Dok2) dan dokumen 3 (Dok3) maka hasil perhitungan ukuran kemiripan untuk contoh di atas dapat dilihat sebagai berikut:

$$\begin{aligned}\text{Sim (Dok2, Q)} &= \frac{(0,352 \times 0,176) + (0,176 \times 0,176) + (0,176 \times 0,176)}{\sqrt{(0,352^2 + 0,176^2 + 0,176^2) \times (0,176^2 + 0,176^2 + 0,176^2)}} \\ &= \frac{0,062 + 0,031 + 0,031}{\sqrt{(0,124 + 0,031 + 0,031) \times (0,031 + 0,031 + 0,031)}} \\ &= \frac{0,124}{\sqrt{(0,186) \times (0,093)}} = \frac{0,124}{\sqrt{0,017}} = \frac{0,124}{0,132} = 0,943\end{aligned}$$

$$\begin{aligned}\text{Sim (Dok3, Q)} &= \frac{(0,176 \times 0,176) + (0,528 \times 0,176) + (0,704 \times 0,176)}{\sqrt{(0,176^2 + 0,528^2 + 0,704^2) \times (0,176^2 + 0,176^2 + 0,176^2)}} \\ &= \frac{0,031 + 0,093 + 0,124}{\sqrt{(0,031 + 0,279 + 0,496) \times (0,031 + 0,031 + 0,031)}} \\ &= \frac{0,248}{\sqrt{(0,806) \times (0,093)}} = \frac{0,248}{\sqrt{0,075}} = \frac{0,248}{0,274} = 0,905\end{aligned}$$

Dari hasil di atas maka dokumen yang memiliki kesamaan dengan *query* “cipta langit bumi” adalah dokumen 2 dan dokumen 3 sehingga jika diurutkan akan terlihat seperti Tabel 4.12 berikut. Hasil dari perangkingan akan disimpan pada tabel kemiripan yang merupakan tabel *temporary*.

Tabel 4.12 Hasil Perangkingan Dokumen Berdasarkan Ukuran Kemiripan Dokumen

Rangking	Dokumen	Kemiripan
I	Dok 2	0,943
II	Dok 3	0,905

4.1.2.3 Seleksi Jawaban (*Answer Selection*)

Tahapan seleksi jawaban adalah tahapan untuk menyeleksi jawaban dari hasil pencarian dokumen yang akan disajikan kepada *user*. Dalam tahapan ini terdapat alur kerja sebagai berikut:

a. Segmentasi Kalimat (*Sentence Segmentation*)

Tahap awal seleksi jawaban adalah segmentasi kalimat. Tiap dokumen hasil pencarian (tahapan *search*) disegmentasi menjadi kalimat-kalimat berdasarkan tanda

titik (.), pada kalimat-kalimat tersebut terlebih dahulu dilakukan *stopword removing* dan *stemming* yang kemudian disimpan pada tabel kalimat. Hasil dari segmentasi kalimat untuk contoh di atas adalah sebagai berikut:

Tabel 4.13 Hasil Segmentasi Kalimat pada Dokumen

Kalimat	Dokumen	Kalimat
Kal2	Dok2	allah cipta langit bumi berkehendak cipta cukup hanya atak kepadanya: jadi lalu jadi
Kal3	Dok3	dalam cipta langit bumi silih ganti malam siang bahtera layar di laut bawa apa guna manusia apa allah turun langit air lalu air itu hidup bumi sudah mati ring bar di bumi itu jenis hewan pengisaran angin awan kendali langit bumi sungguh dapat tanda tanda esa besaran allah kaum memikirkan

b. Jumlah Kata Kunci (*Number of Keywords*)

Setelah disegmentasi kalimat-kalimat yang ada di dokumen, tahap selanjutnya adalah menentukan apakah kalimat-kalimat tersebut memenuhi syarat sebagai kandidat jawaban yang ditentukan dari jumlah kata kunci yang terkandung pada kalimat tersebut. Rumus untuk menghitung jumlah kata kunci dapat dilihat pada persamaan 2.1. Jumlah kata kunci setiap kalimat yang memenuhi syarat disimpan pada tabel *temporary* jumlah_keyword. Hasil dari perhitungan dapat diuraikan sebagai berikut:

Kal2:

Q= 3 (cipt, langit, bumi)

q=4 (cipt, langit, bumi, cipt)

$$q \geq \lfloor \sqrt{Q-1} \rfloor + 1$$

$$4 \geq \lfloor \sqrt{3-1} \rfloor + 1$$

$$4 \geq \lfloor \sqrt{2} \rfloor + 1$$

$$4 \geq \lfloor 1.414 \rfloor + 1$$

4 ≥ 2.414 → cocok, Kal2 memenuhi syarat

Kal3:

Q= 3 (cipt, langit, bumi)

$q=8$ (cipta, langit, bumi, langit, bumi, bumi, langit, bumi)

$$q \geq \lfloor \sqrt{Q-1} \rfloor + 1$$

$$8 \geq \lfloor \sqrt{3-1} \rfloor + 1$$

$$8 \geq \lfloor \sqrt{2} \rfloor + 1$$

$$8 \geq \lfloor 1.414 \rfloor + 1$$

$8 \geq 2.414 \rightarrow$ cocok, Kal3 memenuhi syarat

Dari perhitungan di atas maka didapat bahwa Kal2 dan Kal 3 memenuhi syarat sebagai kandidat jawaban yang akan digunakan pada proses selanjutnya.

c. Penyaring Tipe Jawaban (*Answer Type Filter*)

Seerti yang telah dijelaskan pada bab-bab sebelumnya bahwa Tugas Akhir ini membahas pertanyaan seputar orang, waktu, dan tempat. Pada penyaring tipe jawaban, dibuat sebuah kamus jawaban yang berisi jawaban-jawaban yang terkait dengan pertanyaan mengenai orang, waktu, dan tempat dan ekspresi regular dari pertanyaan dan jawaban yang dibuat secara manual sesuai dari terjemahan Al Quran. Kamus jawaban tersebut disimpan pada tabel `kamus_jawaban`, hasil dari proses ini disimpan pada tabel *temporary* `kandidat_answertype`, dan daftar kamus jawaban dapat dilihat pada Lampiran D.

Tahapan ini mencocokkan pola pertanyaan dengan pola jawaban dan menghitung nilai estimasi untuk setiap pola. Berikut ketentuan yang digunakan pada proses penyaring tipe jawaban yang meliputi tipe jawaban orang, tempat, dan waktu. (Lihat Tabel 4.14)

Tabel 4.14 Ketentuan Tahapan Penyaring Tipe Jawaban

No	Tipe Jawaban	Pola Pertanyaan (nilai estimasi)	Pola Jawaban (nilai estimasi)
1	orang	.*(siapa siapakah).* 1,0	kamus jawaban 0,8 (.*)?(ya para wahai seorang hai) [a-z](.*)? 0,8 (.*)?[a-z] maha(.*)? 0,8 (.*)?di sisi [a-z](.*)? 0,8 (.*)?[a-z](.*)? 0,3
2	tempat	.*(dimana dimanakah).* 1,0 (darimana darimanakah kemana kemanakah).* 1,0	kamus jawaban 0,8 (.*)?(di ke dalam) [a-z](.*)? 0,8 (.*)?ada [a-z](.*)? 0,8 (.*)?[a-z](.*)? 0,3
3	waktu	.*(kapan kapankah).* 1,0	kamus jawaban 0,8 (.*)?(di waktu di hari ke dalam) [a-z](.*)? 0,8 (.*)?[a-z] hari(.*)? 0,8 (.*)?[a-z 0-9](.*)? 0,3

Pada tahap sebelumnya, Kal1 pada Dok2 dan kal2 pada Dok3 memenuhi syarat sehingga kalimat-kalimat tersebut akan diproses pada tahap ini dengan memperhatikan ketentuan-ketentuan pada Tabel 4.14 (Lihat Tabel 4.15).

Tabel 4.15 Hasil Penyaring Tipe Jawaban

Kalimat	Dokumen	Pola Pertanyaan (nilai)	Tipe Jawaban	Pola jawaban	Kandidat Jawaban (nilai)
Kal1	Dok2	orang (1,0)	orang	kamus jawaban	allah (0,8)
				(. *)?[a-z](. *)?	0,3
Kal2	Dok3	orang (1,0)	orang	kamus jawaban	manusia (0,8) allah(0,8) allah(0,8) kaum(0,8)
				(. *)?[a-z](. *)?	0,3

Berdasarkan pada Tabel 4.15 maka nilai dari masing-masing kalimat akan dihitung sesuai tipe (Lihat Tabel 4.16), jika nilai besar dari 0, maka kalimat tersebut akan masuk ke tahap ekstraksi jawaban.

Tabel 4.16 Nilai Kandidat Jawaban

Kalimat	Dokumen	Nilai (nilai estimasi pertanyaan x nilai estimasi jawaban)	Total
Kal1	Dok2	1,0x0,8=0,8 (untuk allah) 1,0x0,3=0,3 (untuk (. *)?[a-z](. *)?)	0,8+0,3=1,1
Kal2	Dok3	1,0x0,3=0,3 (untuk (. *)?[a-z](. *)?) 1,0x0,8=0,8 (untuk manusia) 1,0x0,8=0,8 (untuk allah) 1,0x0,8=0,8 (untuk allah) 1,0x0,8=0,8 (untuk kaum)	0,3+0,8+0,8+0,8+0,8=3,5

Berdasarkan dari Tabel 4.16 tersebut maka Kal1 dan Kal2 akan diproses pada tahap selanjutnya karena nilainya lebih besar dari 0.

d. Ekstraksi Jawaban (*Answer Extraction*) dan *stopword filter*

Tahapan pada ekstraksi jawaban terdiri dari:

1. Mengganti semua kata yang merupakan objek *target* (<TO>) dan *context* (<CO>) pada potongan kalimat yang menjadi kandidat jawaban pada tahapan sebelumnya (*answer type filter*). Objek target (<TO>) dan *context* (<CO>) di dapat dari tahap *query generator* dalam pembentukan *query* (*Query Formation*). Objek target (<TO>) untuk kasus ini adalah cipta, dan objek *context*-nya (<CO>) adalah langit bumi. Hasil dari tahap awal ini dapat dilihat pada Tabel 4.17.

Tabel 4.17 Hasil Ekstraksi <TO> dan <CO>

Kalimat	Dokumen	Hasil
Kal1	Dok2	allah <TO> <CO> berkehendak <TO> cukup hanya atak kepadanya: jadi lalu jadi
Kal2	Dok3	dalam <TO> <CO> silih ganti malam siang bahtera layar di laut bawa apa guna manusia apa allah turun langit air lalu air itu hidup bumi sudah mati ring bar di bumi itu jenis hewan pengisaran angin awan kendali <CO> sungguh dapat tanda tanda esa besaran allah kaum memikirkan

2. Tahap selanjutnya adalah mengekstrak objek *property* <PO> yang sesuai dengan pola jawaban yang telah dihitung nilai *confidence* dan *support*-nya pada tahap pengujian pola jawaban (Lihat subbab 5.2.3.1 halaman 5-10) yang dapat dilihat pada Tabel 4.18. Untuk hasil untuk kasus di atas dapat dilihat pada Tabel 4.19.

Tabel 4.18 Pola Jawaban (*Answer Patterns*)

No	Property	Pola Jawaban	Benar	Salah	Confidence	Support
1	orang	<PO_orang> ([a-z]) <TO>	1	0	1	0,0023
2	orang	<PO_orang> <TO>	61	73	0,455224	0,1422
3	orang	wahai <PO_orang> <TO>	2	0	1	0,0047
4	orang	<PO_orang> (. *) ? <TO>	135	462	0,226131	0,3147
5	orang	<TO> (. *) ? <PO_orang>	131	530	0,198185	0,3054

Tabel 4.18 Pola Jawaban (*Answer Patterns*) (Lanjutan)

No	Property	Pola Jawaban	Benar	Salah	Confidence	Support
6	orang	<TO> <PO_orang>	20	37	0,350877	0,0466
7	orang	<PO_orang> di waktu <TO>	2	0	1	0,0047
8	orang	ialah <PO_orang> <TO>	4	1	0,8	0,0093
9	orang	apa <TO> <PO_orang>	2	10	0,166667	0,0047
10	orang	ya <PO_orang> <TO>	2	3	0,4	0,0047
11	tempat	<PO_tempat> itu tempat <TO>	3	0	1	0,0171
12	tempat	<TO> ialah <PO_tempat>	1	0	1	0,0057
13	tempat	ialah <PO_tempat>(.*)?<TO>	1	0	1	0,0057
14	tempat	sungguh <PO_tempat> itu tempat <TO>	1	0	1	0,0057
15	waktu	<TO>(.*)?<PO_waktu>	24	43	0,358209	0,4000
16	waktu	<PO_waktu>(.*)?<TO>	10	39	0,204082	0,1667
17	waktu	<TO>(.*)?di <PO_waktu>	3	10	0,230769	0,0500
18	orang	<TO> selain <PO_orang>	6	4	0,6	0,0140
19	orang	ketika <PO_orang>(.*)?<TO>	1	7	0,125	0,0023
20	orang	hai <PO_orang> <TO>	2	1	0,666667	0,0047
21	orang	<PO_orang> itu <TO>	1	4	0,2	0,0023
22	tempat	<PO_tempat>(.*)?<TO>	38	126	0,231707	0,2171
23	tempat	<TO>(.*)?di <PO_tempat>	2	11	0,153846	0,0114
24	tempat	itu <PO_tempat> tempat <TO>	1	0	1	0,0057
25	tempat	<TO> <PO_tempat>	7	19	0,269231	0,0400
26	tempat	<PO_tempat> adalah <TO>	1	0	1	0,0057
27	tempat	<TO> ke <PO_tempat>	3	4	0,428571	0,0171
28	tempat	<TO>(.*)?<PO_tempat>	33	138	0,192982	0,1886
29	tempat	tempat <TO>(.*)?<PO_tempat>	2	0	1	0,0114
30	tempat	<TO>(.*)?ke dalam <PO_tempat>	4	26	0,133333	0,0229
31	waktu	<PO_waktu>(.*)?dalam <TO>	1	0	1	0,0167
32	waktu	<TO> <PO_waktu>	8	16	0,333333	0,1333
33	waktu	<TO>(.*)?lama <PO_waktu>	1	0	1	0,0167
34	waktu	di <PO_waktu>(.*)?<TO>	4	7	0,363636	0,0667

Dari Tabel 4.18 di atas maka Kal1 sesuai dengan pola jawaban <PO_orang> <TO> dan <PO_orang>(.*)?<TO>. Kal2 sesuai dengan pola jawaban <TO>(.*)?<PO_orang>, maka hasil ekstraksi dan *stopword filter*-nya dapat dilihat pada Tabel 4.19.

Tabel 4.19 Hasil Ekstraksi <PO>

Kalimat	Dokumen	Pola Jawaban	Hasil
Kal1	Dok2	<PO_orang> <TO>	allah (0,455224)
		<PO_orang>(.*)?<TO>	allah (0,226131)
Kal2	Dok3	<TO>(.*)?<PO_orang>	allah (0,198185) manusia (0,198185) kaum (0,198185)

Hasil ekstraksi pada Tabel 4.19 kemudian disimpan pada tabel *temporary* jawaban, dan untuk kandidat jawaban yang sama yang terdapat pada kalimat yang sama maka nilai *confidence*-nya dijumlahkan sesuai dengan pola jawaban. Hasilnya diurutkan berdasarkan nilai tertinggi yang dapat dilihat pada Tabel 4.20 berikut:

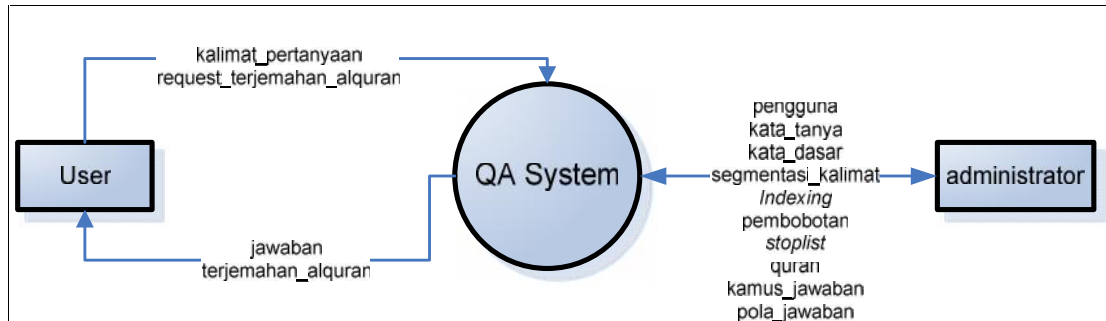
Tabel 4.20 Kandidat Jawaban

No	Kalimat	Kandidat Jawaban	Nilai
1	Kal1	Allah	0,681355
2	Kal2	Allah	0,198185
3	Kal2	Manusia	0,198185
4	Kal2	Kaum	0,198185

Berdasarkan Tabel 4.20, maka jawaban yang akan diberikan kepada *user* adalah allah di peringkat pertama dan kedua, manusia dan kaum pada peringkat 3 dan 4.

4.1.3 Analisa Fungsional Sistem

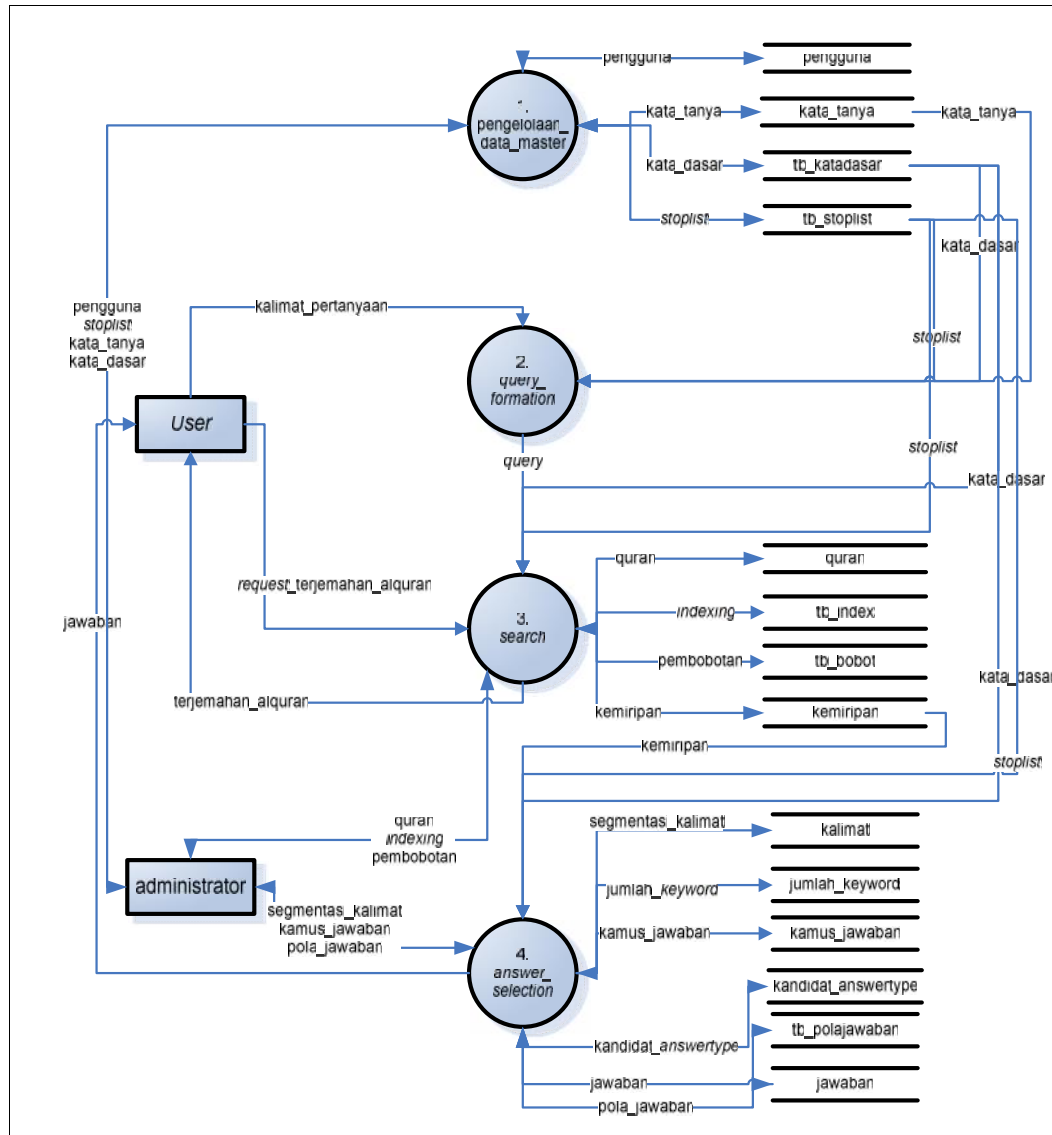
Pada analisa fungsional akan dijelaskan mengenai *context diagram* yang dapat dilihat pada Gambar 4.4 dan keterangannya dapat dilihat pada Tabel 4.21 dan *Data Flow Diagram level 1* dapat dilihat pada Gambar 4.5 dan keterangannya dapat dilihat pada Tabel 4.22. Untuk penguraian *Data Flow Diagram level 2* dan seterusnya dapat dilihat pada Lampiran E.



Gambar 4.4. *Context Diagram* Aplikasi QA System

Tabel 4.21 Tabel *Context Diagram* Aplikasi QA System

No	Jenis	Nama	Deskripsi
1.	Entitas	<i>Administrator</i>	<i>Administrator</i> berperan sebagai pengelola sistem
		<i>User</i>	Pengguna sistem
2.	Aliran Data	kalimat_pertanyaan	Kalimat pertanyaan yang di-inputkan oleh <i>user</i>
		Jawaban	Jawaban yang dikembalikan sistem untuk menjawab pertanyaan yang di-inputkan <i>user</i>
		Pengguna	Data pengguna
		kata_tanya	Data kata tanya
		kata_dasar	Daftar kata dasar bahasa Indonesia
		segmentasi_kalimat	Berisi data hasil segmentasi kalimat
		<i>Indexing</i>	Berisi data hasil <i>indexing</i>
		Pembobotan	Berisi data pembobotan
		<i>Stoplist</i>	Berisi kata-kata yang dianggap sebagai <i>stopwords</i>
		quran terjemahan_alquran request_terjemahan_alquran	Berisi terjemahan ayat-ayat Al Quran
		kamus_jawaban	Berisi jawaban-jawaban yang mungkin yang terdapat dalam terjemahan Al Quran
		pola_jawaban	Data pola jawaban



Gambar 4.5. DFD Level 1 Aplikasi QA System

Tabel 4.22 Tabel DFD Level 1 Aplikasi QA System

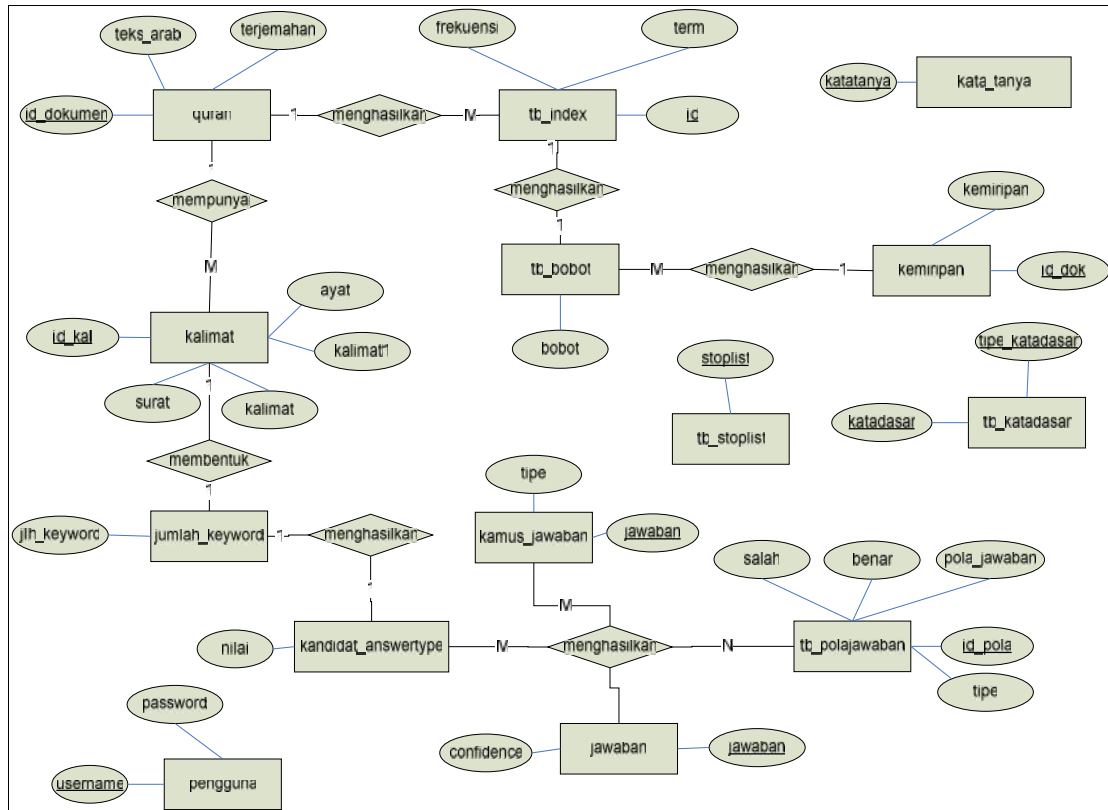
No	Jenis	Nama	Deskripsi
1.	Proses	pengelolaan data master	Pengelolaan data master
		<i>query_formation</i>	Proses yang melakukan analisa terhadap pertanyaan untuk menghasilkan <i>query</i>
		<i>Search</i>	Proses yang melakukan pencarian dokumen yang

Tabel 4.22 Tabel DFD Level 1 Aplikasi QA System (Lanjutan)

No	Jenis	Nama	Deskripsi
			relevan dengan <i>query</i>
		<i>answer_selection</i>	Proses untuk menyeleksi jawaban yang tepat
2.	Aliran Data	kalimat_pertanyaan	Kalimat pertanyaan yang diinputkan oleh <i>user</i>
		Jawaban	Jawaban yang dikembalikan sistem untuk menjawab pertanyaan yang di- <i>input</i> -kan <i>user</i>
		Pengguna	Berisi data pengguna
		kata_dasar	Daftar kata dasar dalam bahasa Indonesia
		kata_tanya	Data kata tanya
		<i>Query</i>	Berisi <i>query</i> hasil proses <i>query formation</i>
		Kemiripan	Berisi nilai <i>similarity</i> dari dokumen
		<i>Stoplist</i>	Berisi kata-kata yang dianggap sebagai <i>stopwords</i>
		<i>Indexing</i>	Berisi data hasil <i>indexing</i>
		Pembobotan	Berisi data pembobotan
		quran terjemahan_alquran <i>request_terjemahan_alquran</i>	Berisi terjemahan ayat-ayat Al Quran
		segmentasi_kalimat	Berisi data hasil segmentasi kalimat
		jumlah_keyword	Berisi jumlah kata kunci yang terdapat pada kalimat di dalam dokumen
		kandidat_ <i>answertype</i>	Berisi data hasil proses <i>answer type filter</i>
		kamus_jawaban	Berisi jawaban-jawaban yang mungkin yang terdapat dalam terjemahan Al Quran
		pola_jawaban	Data pola jawaban

4.1.4 Analisa Data Sistem

Pada analisa data sistem akan dijelaskan mengenai *Entity Relation Diagram* (ERD) yang dapat dilihat pada Gambar 4.6 dan keterangannya dapat dilihat pada Tabel 4.23.



Gambar 4.6. ERD Aplikasi QA System

Tabel 4.23 Tabel ERD QA System

N o	Jenis	Nama	Deskripsi	Atribut	Primary Key
1	Entitas	tb_stoplist	Menyimpan data kata yang termasuk dalam <i>stopwords</i>	stoplist	stoplist
		jumlah_key- word	Menyimpan data jumlah kata kunci yang terdapat dalam kalimat pada dokumen	jlh_keyword	
		kamus_jawa- ban	Menyimpan data jawaban yang memungkinkan yang terdapat pada terjemahan Al Quran	jawaban tipe	jawaban
		kandidat_ answertype	Menyimpan data hasil proses <i>answer type filter</i>	nilai	
		kata_tanya	Menyimpan data kata tanya	katatanya	

Tabel 4.23 Tabel ERD QA System (Lanjutan)

N o	Jenis	Nama	Deskripsi	Atribut	Primary Key
		kemiripan	Menyimpan data kemiripan dokumen	id_dok kemiripan	id_dok
		pengguna	Menyimpan data pengguna	<i>username</i> <i>password</i>	<i>username</i>
		quran	Menyimpan data quran dan terjemahannya	id_dokumen teks_arab terjemahan	id_dokumen
		kalimat	Menyimpan data kalimat hasil segmentasi kalimat	id_kal kalimat kalimat1 surat ayat	id_kal
		tb_polajawaban	Menyimpan data pola jawaban	id_pola pola_jawaban tipe salah benar	id_pola
		tb_katadasar	Menyimpan data kata dasar dalam bahasa Indonesia	katadasar tipe_katadasar	katadasar tipe_katadasar
		tb_index	Menyimpan data hasil pengindeksan	id <i>term</i> frekuensi	id
		tb_bobot	Menyimpan data hasil pembobotan	bobot	
		jawaban	Menyimpan data jawaban yang akan ditampilkan kepada <i>user</i>	jawaban <i>confidence</i>	jawaban
	2	Relasi	mempunyai		
		menghasilkan	Hubungan antara entitas tb_index dengan quran, entitas tb_index dengan tb_bobot, entitas tb_bobot dengan kemiringan, entitas jumlah_keyword dengan kandidat_answertype, dan entitas		

Tabel 4.23 Tabel ERD QA System (Lanjutan)

No	Jenis	Nama	Deskripsi	Atribut	Primary Key
			kandidat_answertype, kamus_jawaban, tb_polajawaban dengan jawaban		

4.2.Perancangan Sistem

Perancangan sistem dikelompokkan menjadi dua sub pokok bahasan, yaitu perancangan tabel, perancangan menu dan perancangan antarmuka.

4.2.1 Perancangan Tabel

- a. Tabel pengguna, digunakan untuk menampung data *administrator*.

Tabel 4.24 Tabel pengguna

Nama Field	Type dan Length	Primary Key	Null	Default
<i>Username</i>	Varchar(20)	Yes	No	-
<i>Password</i>	Varchar(50)		No	-

- b. Tabel *tb_stoplist*, digunakan untuk menampung kata-kata yang termasuk dalam *stopwords*.

Tabel 4.25 Tabel *tb_stoplist*

Nama Field	Type dan Length	Primary Key	Null	Default
<i>stoplist</i>	Varchar(20)	Yes	No	-

- c. Tabel *tb_katadasar*, digunakan untuk menampung kata dasar bahasa Indonesia

Tabel 4.26 Tabel *tb_katadasar*

Nama Field	Type dan Length	Primary Key	Null	Default
katadasar	Varchar(20)	Yes	No	-
tipe_katadasar	Varchar(10)	Yes	No	-

- d. Tabel quran, digunakan untuk menampung data terjemahan Al Quran.

Tabel 4.27 Tabel quran

Nama Field	Type dan Length	Primary Key	Null	Default
id_dokumen	Varchar(5)	Yes	No	-
teks_arab	Text		No	-
terjemahan	Varchar(12)		No	-

- e. Tabel kemiripan, digunakan untuk menampung data sementara kemiripan dokumen.

Tabel 4.28 Tabel kemiripan

Nama Field	Type dan Length	Primary Key	Null	Default
id_dok	Varchar(5)	Yes	No	-
kemiripan	Float		No	-

- f. Tabel kamus_jawaban, digunakan untuk menampung data jawaban-jawaban yang memungkinkan pada terjemahan Al Quran.

Tabel 4.29 Tabel kamus_jawaban

Nama Field	Type dan Length	Primary Key	Null	Default
jawaban	Varchar(100)	Yes	No	-
tipe	Varchar(7)		No	-

- g. Tabel kalimat, digunakan untuk menampung data kalimat hasil segmentasi pada dokumen.

Tabel 4.30 Tabel kalimat

Nama Field	Type dan Length	Primary Key	Null	Default
id_kal	Integer(5)	Yes	No	-
id_dok	Varchar(5)		No	-
kalimat	Text		No	-
kalimat1	Text		No	-
surat	Varchar(20)		No	-
ayat	Varchar(15)		No	-

- h. Tabel *kata_tanya*, digunakan untuk menampung data kata tanya.

Tabel 4.31 Tabel *kata_tanya*

Nama Field	Type dan Length	Primary Key	Null	Default
katatanya	Varchar(15)	Yes	No	-

- i. Tabel *jumlah_keyword*, digunakan untuk menampung data jumlah kata kunci pada kalimat di dalam dokumen.

Tabel 4.32 Tabel *jumlah_keyword*

Nama Field	Type dan Length	Primary Key	Null	Default
id_kal	Integer(5)	Yes	No	-
jlh_keyword	Integer(4)		No	-

- j. Tabel *jawaban*, digunakan untuk menampung data sementara kandidat jawaban.

Tabel 4.33 Tabel *jawaban*

Nama Field	Type dan Length	Primary Key	Null	Default
id_pola	Integer(4)	Yes	No	-
id_kal	Integer(5)	Yes	No	-
jawaban	Varchar(500)	Yes	No	-
<i>confidence</i>	Float		No	-

- k. Tabel *kandidat_answer_type*, digunakan untuk menampung data sementara hasil proses *answer type filter*.

Tabel 4.34 Tabel *kandidat_answer_type*

Nama Field	Type dan Length	Primary Key	Null	Default
id_kal	Integer(5)	Yes	No	-
nilai	Float		No	-

- l. Tabel *index*, digunakan untuk menampung data hasil pengindeksan.

Tabel 4.35 Tabel *index*

Nama Field	Type dan Length	Primary Key	Null	Default
id	Integer(10)	Yes	No	-
<i>term</i>	Varchar(30)		No	-
dok_id	Varchar(5)		No	-
frekuensi	Integer(11)		No	-

- m. Tabel *tb_polajawaban*, digunakan untuk menampung data pola jawaban.

Tabel 4.36 Tabel pola jawaban

Nama Field	Type dan Length	Primary Key	Null	Default
<i>id_pola</i>	Integer(4)	Yes	No	-
<i>Tipe</i>	Varchar(7)		No	-
<i>pola_jawaban</i>	Varchar(100)		No	-
<i>benar</i>	Integer(4)		No	0
<i>Salah</i>	Integer(4)		No	0

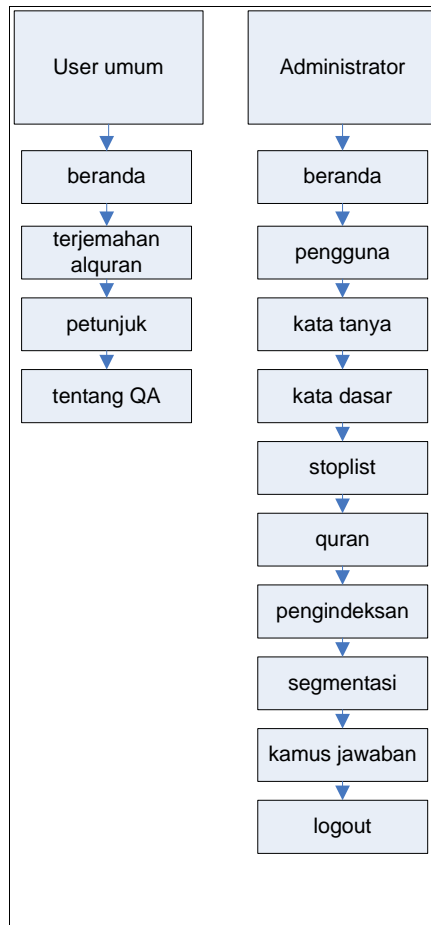
- n. Tabel *bobot*, digunakan untuk menampung data hasil pembobotan.

Tabel 4.37 Tabel bobot

Nama Field	Type dan Length	Primary Key	Null	Default
<i>Term</i>	Varchar(30)	Yes	No	-
<i>dok_id</i>	Varchar(5)	Yes	No	-
<i>bobot</i>	Float		No	-

4.2.2 Perancangan Menu

Berikut adalah perancangan struktur menu dari sistem yang dirancang agar memudahkan di dalam melakukan integrasi antar modul. Sistem memiliki dua *user* di mana *privilege* dari masing-masing *user* berbeda sesuai dengan fungsi masing-masing. Untuk tiap *user* memiliki struktur menu yang berbeda. Struktur menu dapat dilihat pada Gambar 4.7.



Gambar 4.7. Struktur Menu Aplikasi QA System

4.2.3 Perancangan Antarmuka

Pada sub bab ini, hanya menjelaskan rancangan antarmuka pada proses-proses utama dari aplikasi QA System untuk terjemahan Al Quran *form* utama, dan hasil pencarian. Antarmuka yang dibangun adalah sebagai berikut:

4.2.3.1 Tampilan Utama

Tampilan pada halaman utama ini adalah tampilan utama dari aplikasi saat dieksekusi. Pada tampilan utama ini terdapat sebuah *form* pencarian yang akan digunakan oleh pengguna untuk memasukkan kalimat pertanyaan.

(Lihat Gambar 4.8)

Gambar 4.8. Antarmuka *Form* Utama

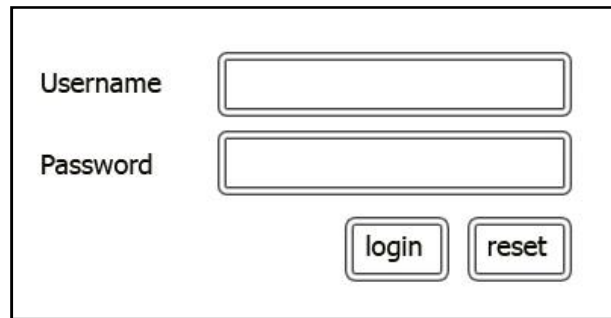
4.2.3.2 Tampilan Hasil Pencarian

Pada *form* ini terdapat sebuah *form* pencarian untuk memasukkan kalimat pertanyaan dan terdapat daftar jawaban dari pertanyaan yang dikembalikan oleh QA *System*. (Lihat Gambar 4.9)

Gambar 4.9. Antarmuka *Form* Hasil Pencarian

4.2.3.3 Form Login Untuk Administrator

Pada aplikasi QA System ini terdapat halaman khusus untuk *administrator* yang bertujuan untuk mengelola sumber pengetahuan sistem yang dilakukan secara manual. Untuk mengakses halaman *administrator* harus terlebih dahulu melakukan *login*. Antarmuka *Form login* digambarkan pada Gambar 4.10.

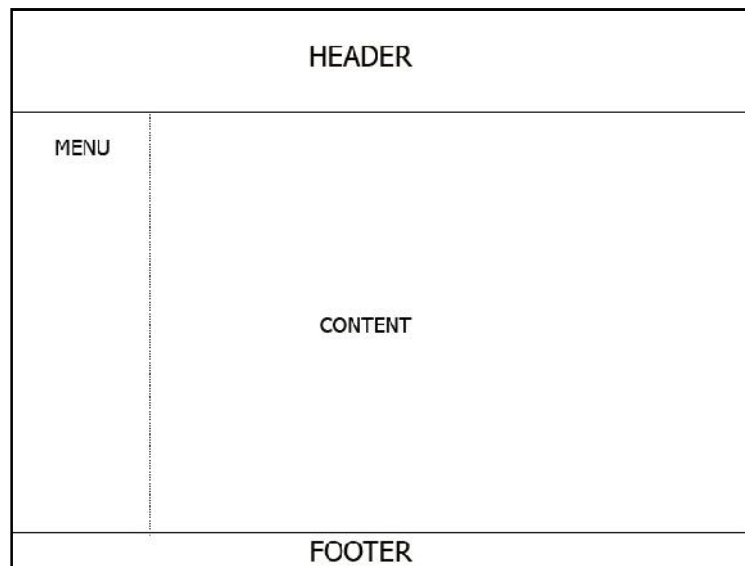


The image shows a login form for an administrator. It consists of a rectangular box containing two input fields. The first field is labeled 'Username' and the second is labeled 'Password'. Below these fields are two buttons: 'login' and 'reset'.

Gambar 4.10. Antarmuka *Form Login* untuk *Administrator*

4.2.3.4 Halaman Utama Untuk Administrator

Antarmuka pada Gambar 4.11 di bawah ini merupakan antarmuka yang akan digunakan untuk setiap menu yang terdapat dalam halaman *administrator*.



The image shows a layout for the main page of an administrator. It is a rectangular box divided into four sections: 'HEADER' at the top, 'FOOTER' at the bottom, 'MENU' on the left side, and 'CONTENT' in the center.

Gambar 4.11. Antarmuka Halaman Utama untuk *Administrator*

BAB V

IMPLEMENTASI DAN PENGUJIAN

5.1. Implementasi Sistem

Pada bab ini implementasi perangkat lunak akan diuraikan sesuai dengan analisa dan perancangan yang telah dijelaskan pada bab sebelumnya.

5.1.1 Batasan Implementasi

Batasan implementasi pada Tugas Akhir ini adalah sebagai berikut:

1. Aplikasi ini dibuat berbasis web dengan menggunakan bahasa pemrograman PHP dan basis data MySQL.
2. Data yang digunakan adalah terjemahan alquran dari Departemen Agama revisi terjemah tahun 1989.

5.1.2 Lingkungan Implementasi

Lingkungan implementasi perangkat keras dan perangkat lunak pada perancangan aplikasi ini adalah sebagai berikut:

1. Perangkat Keras
 - a. *Processor* : *Intel(R) Core(TM) i5-2450M CPU @ 2.50GHz*
 - b. *Memory* : 8.00 GB
2. Perangkat Lunak
 - a. Sistem Operasi : *Windows 7 Home Premium 64-bit (6.1,Build 7601)*
 - b. Bahasa Pemrograman : *PHP Script Language Version 5.2.6.*
Tools perancangan : Notepad ++ versi 5.9.3.
Browser : Google Chrome, Mozilla FireFox
DBMS : MySQL

Web server : Apache Web Server Version 2.2.8.

5.1.3 Hasil Implementasi

Implementasi rancang bangun dari QA System yang dijelaskan pada bab ini adalah implementasi pada proses-proses utama sedangkan untuk proses-proses pendukung dapat dilihat pada Lampiran F.





1. Menu Utama

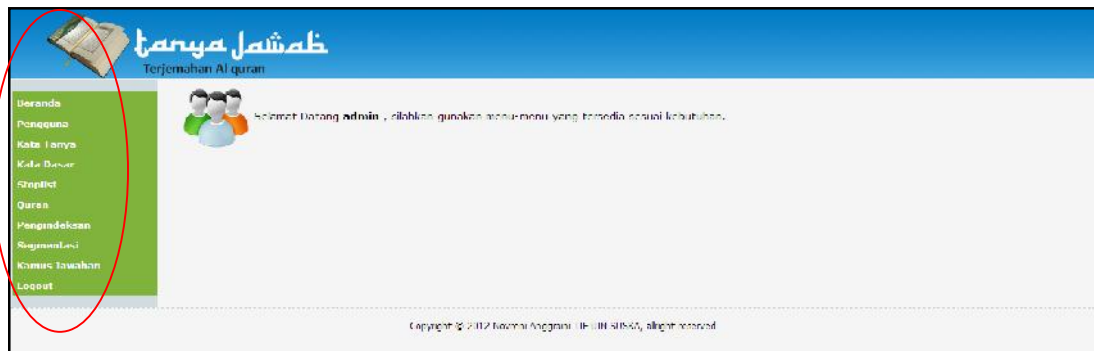
Aplikasi QA System ini memiliki tampilan *front end* dan *back end*. Tampilan *front end* dapat diakses oleh user pada umumnya, dan tampilan *back end* hanya boleh diakses oleh administrator. Pada sisi *front end* terdapat empat menu utama yaitu beranda, terjemahan Al Quran, petunjuk, dan tentang QA (Lihat Gambar 5.1 dan Tabel 5.1). Sisi *back end* memiliki sebelas menu utama yaitu beranda, pengguna, kata tanya, kata dasar, *stoplist*, quran, pengindeksan, segmentasi, kamus jawaban dan *logout* (Lihat Gambar 5.2 dan Tabel 5.2).



Gambar 5.1. Menu Utama *Front End* Aplikasi QA System

Tabel 5.1 Deskripsi Menu Utama *Front End* Aplikasi QA System

Menu	Ikon	Deskripsi
Beranda		Merupakan menu utama yang digunakan untuk memasukkan pertanyaan.
Terjemahan Al Quran		Merupakan menu yang digunakan untuk melihat terjemahan Al Quran secara keseluruhan.
Petunjuk		Merupakan menu yang digunakan sebagai petunjuk penggunaan aplikasi.
Tentang QA		Merupakan menu yang digunakan untuk melihat spesifikasi dari aplikasi.



Gambar 5.2. Menu Utama *Back End* Aplikasi QA System

Tabel 5.2 Deskripsi Menu Utama *Back End* Aplikasi QA System

Menu	Deskripsi
Beranda	Merupakan menu utama yang digunakan untuk mengakses halaman utama aplikasi.
Pengguna	Merupakan menu yang digunakan untuk mengelola data pengguna.
Kata Tanya	Merupakan menu yang digunakan untuk mengelola data kata tanya.
Kata Dasar	Merupakan menu yang digunakan untuk mengelola data kata dasar.
<i>Stoplist</i>	Merupakan menu yang digunakan untuk mengelola data <i>stoplist</i> .
Quran	Merupakan menu yang digunakan untuk mengelola data Al Quran.
Pengindeksan	Merupakan menu yang digunakan untuk mengelola <i>indexing</i> yang digunakan pada tahapan pencarian dokumen.
Segmentasi	Merupakan menu yang digunakan untuk mengelola proses segmentasi kalimat pada tahapan ekstraksi jawaban.
Kamus Jawaban	Merupakan menu yang digunakan untuk mengelola data kamus jawaban.
<i>Logout</i>	Merupakan menu yang digunakan untuk keluar dari aplikasi.

2. Form Tanya Jawab

Form tanya jawab merupakan sebuah *form* yang digunakan untuk memasukkan pertanyaan yang akan ditanyakan. *Form* ini terletak pada halaman *front end* yang dapat diakses oleh *user* pada umumnya. Tampilan *form* tanya jawab dapat dilihat pada Gambar 5.3. Setelah tombol tanya pada *form* tersebut di klik, maka

aplikasi akan menampilkan daftar jawaban dari pertanyaan (Lihat Gambar 5.4) dan jawaban-jawaban tersebut dapat dilihat secara rinci seperti Gambar 5.5.



The screenshot shows the 'Form Tanya Jawab' (Question Answer Form) of the QA System. The interface has a green header with the logo 'ظم شامل عيبره' (Zam Shamil Ayyirah) and 'Terjemahan Al Quran'. Below the header, there is a search bar labeled 'Ketikkan Pertanyaan' (Enter Question) with the text 'Siapa pencipta langit dan bumi?' (Who is the creator of the sky and earth?). To the right of the search bar is a green button labeled 'Tanya' (Ask). Above the search bar, there are four icons: a home icon, a folder icon, a question mark icon, and an information icon.

Gambar 5.3. *Form Tanya Jawab Aplikasi QA System*



The screenshot shows the 'Daftar Jawaban' (List of Answers) of the QA System. The interface is the same as in Gambar 5.3, but now it displays the answers to the question 'Siapa pencipta langit dan bumi?'. Below the search bar, there is a green button labeled 'Tanya' (Ask). Below the button, there is a message: 'Ditemukan daftar dokumen sebanyak 83 yang mengandung jawaban.' (Found a list of documents as many as 83 containing the answer.). Below the message, there are two answers listed:

- 1 Jawab : Allah**
katakanlah: wahai **allah** pencipta langit dan bumi yang mengetahui barang ghaib dan yang nyata engkaulah yang membolak-balikan antara hamba-hamba-mu (antara) apa yang selalu mereka mongoteksifikannya ... (QS.al-Qamar:46)
selengkapnya
- 2 Jawab : Allah**
sesungguhnya tuhan kamu ialah **allah** yang menciptakan langit dan bumi dalam enam masa kemudian dia bergeayam di atas arsy untuk mengatur segala urusan ... (QS.yunus:3)
selengkapnya

Gambar 5.4. *Daftar Jawaban Aplikasi QA System*



Gambar 5.5. Detil Jawaban Aplikasi QA System

3. Menu Terjemahan Al Quran

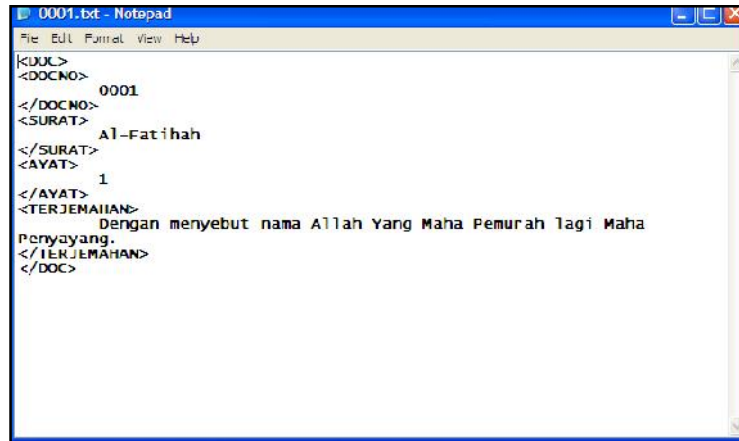
Menu terjemahan Al Quran adalah sebuah menu yang berfungsi untuk menampilkan terjemahan Al Quran yang terletak pada halaman *front end*. Terjemahan Al Quran ditampilkan berdasarkan urutan nama surat dari Al Quran. Tampilan terjemahan Al Quran dapat dilihat pada Gambar 5.6 berikut.



Gambar 5.6. Tampilan Terjemahan Al Quran Aplikasi QA System

4. Korpus Terjemahan Al Quran

Pada Tugas Akhir ini korpus dibuat dengan format teks (*.txt) yang mengandung *tag-tag* XML yang dapat dilihat pada Gambar 5.7.



Gambar 5.7. Tampilan Korpus Terjemahan Al Quran Aplikasi QA System

5. Stoplist

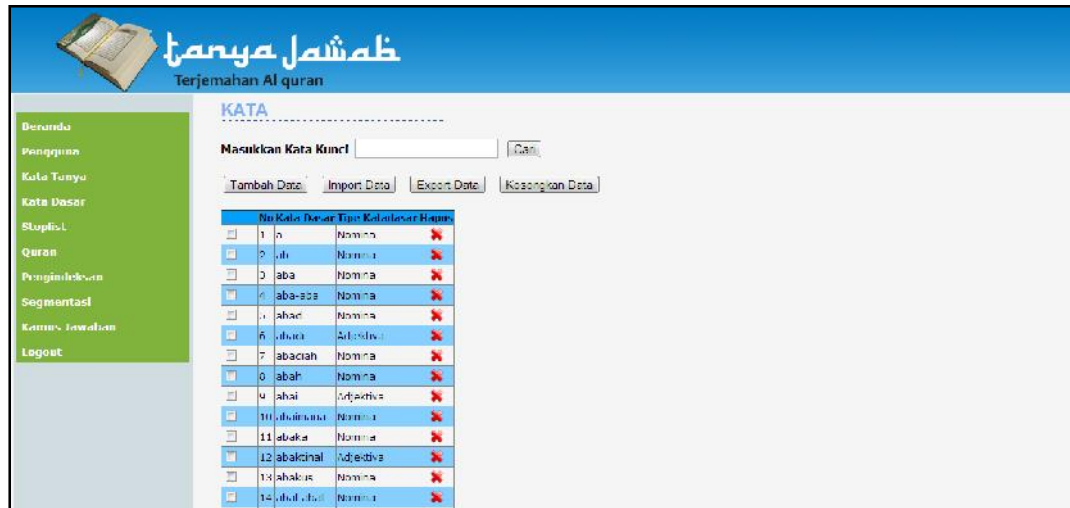
Stoplist pada perancangan digunakan pada proses *stopword removal* yang disimpan dalam *database*. *Stoplist* terletak pada halaman *back end* yang hanya bisa diakses oleh administrator. Tampilan halaman *stoplist* dapat dilihat pada Gambar 5.8.



Gambar 5.8. Tampilan Halaman *Stoplist* Aplikasi QA System

6. Kata Dasar

Kata dasar digunakan pada proses *stemming*, kata dasar disimpan dalam *database* dan terletak pada halaman *back end*. Tampilan halaman kata dasar dapat dilihat pada Gambar 5.9.



Gambar 5.9. Tampilan Halaman Kata Dasar Aplikasi QA System

7. Pengindeksan

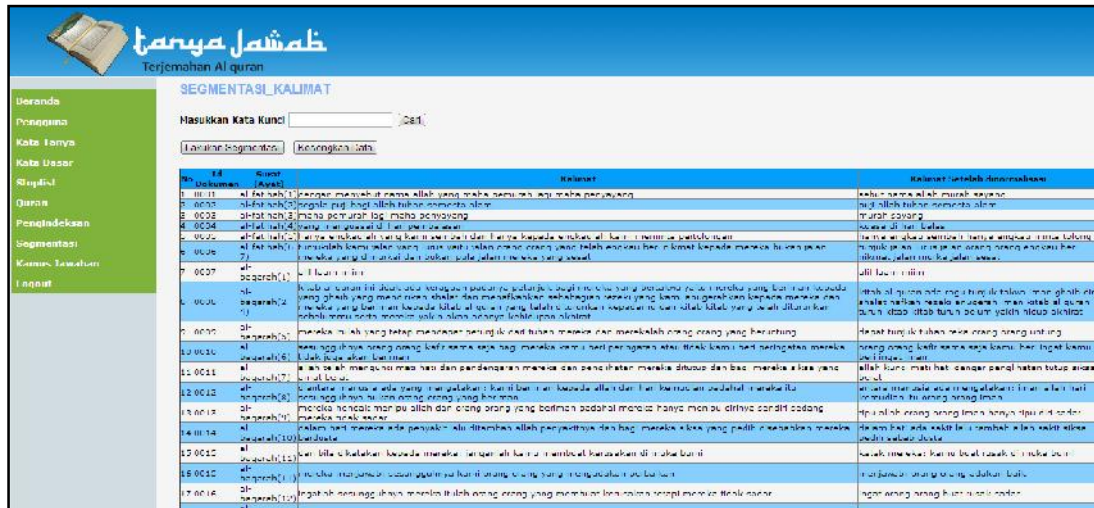
Pengindeksan digunakan pada proses pencarian (*information retrieval*), pengindeksan disimpan dalam *database* dan terletak pada halaman *back end*. Tampilan halaman pengindeksan dapat dilihat pada Gambar 5.10.



Gambar 5.10. Tampilan Halaman Pengindeksan Aplikasi QA System

8. Segmentasi

Segmentasi digunakan pada proses ekstraksi jawaban. Segmentasi disimpan dalam *database* dan terletak pada halaman *back end*. Tampilan halaman segmentasi dapat dilihat pada Gambar 5.11.



No	Kata Kunci	Jawaban	Jawaban setelah diinterpretasikan
1	al-fahrah(1)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
2	al-fahrah(2)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
3	al-fahrah(3)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
4	al-fahrah(4)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
5	al-fahrah(5)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
6	al-fahrah(6)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
7	al-fahrah(7)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
8	al-fahrah(8)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
9	al-fahrah(9)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
10	al-fahrah(10)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
11	al-fahrah(11)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
12	al-fahrah(12)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
13	al-fahrah(13)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
14	al-fahrah(14)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
15	al-fahrah(15)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
16	al-fahrah(16)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
17	al-fahrah(17)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling
18	al-fahrah(18)	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling	pengantar merupakan nama Allah yang telah diciptakan oleh Allah yang paling

Gambar 5.11. Tampilan Halaman Segmentasi Aplikasi QA System

9. Kamus Jawaban

Kamus jawaban merupakan sumber pengetahuan (*knowledge source*) yang digunakan pada tahap penyaringan tipe jawaban (*answer type filter*) dan ekstraksi jawaban (*answer extraction*). Kamus jawaban hanya dapat diakses oleh administrator. Tampilan halaman ini dapat dilihat pada Gambar 5.12.



Gambar 5.12. Tampilan Halaman Kamus Dasar Aplikasi QA System

5.2. Pengujian

Dalam pengujian rancang bangun aplikasi QA System ini terdapat beberapa lingkungan pengujian, hasil pengujian, dan kesimpulan pengujian. Berikut akan dijelaskan masing-masing pembahasan.

5.2.1 Lingkungan Pengujian

1. Perangkat Keras

- a. *Processor* : Intel(R) Core(TM) i5-2450M CPU @ 2.50GHz
- b. *Memory* : 8.00 GB

2. Perangkat Lunak

- a. *Sistem Operasi* : Windows 7 Home Premium 64-bit (6.1, Build 7601)
- b. *Bahasa Pemrograman* : PHP Script Language Version 5.2.6.

Tools perancangan : Notepad ++ versi 5.9.3.

Browser : Google Chrome, Mozilla FireFox

DBMS : MySQL

Web server : Apache Web Server Version 2.2.8.

5.2.2 Pengujian Sistem

Pengujian sistem ini dilakukan pada 5225 *file text* (1,62 MB) yang terdiri dari 114 surat pada terjemahan Al Quran. Pada pengujian ini, terdapat beberapa aspek yang diuji, yaitu pengujian terhadap pola jawaban (*answer patterns*), pengujian terhadap aspek presisi (*precision*), CWS (*Confidence Wighted Score*) serta pengujian terhadap interpretasi pertanyaan.

5.2.2.1 Pengujian terhadap Pola Jawaban (*Answer Patterns*)

Pengujian ini berfungsi sebagai data latih untuk menyaring pola-pola jawaban yang akan digunakan sebagai data pola jawaban (*answer patterns*) untuk mengekstrak jawaban dari pertanyaan yang diajukan *user* (Lihat bab IV). Pengujian terhadap pola jawaban (*answer patterns*) aplikasi ini dilakukan dengan menguji 45 pertanyaan berdasarkan kategori orang, waktu, dan tempat dimana pertanyaan untuk kategori orang terdiri dari 15 pertanyaan, tempat 20 pertanyaan, dan 10 pertanyaan untuk kategori waktu. Pada pengujian ini diasumsikan bahwa tiap pertanyaan memiliki minimal satu buah jawaban. Pertanyaan-pertanyaan tersebut dapat dilihat pada Tabel 5.3.

Tabel 5.3 Pertanyaan yang diujikan untuk Pola Jawaban (*Answer Patterns*)

No	Kategori	Pertanyaan
1	orang	Siapakah pencipta langit dan bumi?
2	orang	Siapakah yang ingkar kepada ayat-ayat allah?
3	orang	Siapakah yang harus disembah manusia?
4	orang	Siapakah yang menyuruh kamu menyembelih seekor sapi betina?
5	orang	Siapakah yang memberi rezeki?
6	orang	Kepunyaan siapakah apa yang di langit dan di bumi?
7	orang	Maqam siapakah yang dijadikan tempat shalat?
8	orang	Diwahyukan kepada siapakah al quran?
9	orang	Putera siapakah isa?
10	orang	Dibunuh oleh siapakah jalut?
11	orang	Taurat diberikan kepada siapa?
12	orang	Puasa diwajibkan kepada siapa?
13	orang	Neraka disediakan untuk siapa?
14	orang	Matahari diterbitkan oleh siapa?
15	orang	Surga dihuni oleh siapa?

Tabel 5.3 Pertanyaan yang diujikan untuk Pola Jawaban (*Answer Patterns*)

(Lanjutan)

No	Kategori	Pertanyaan
16	tempat	Dimanakah tempat berkumpul bagi manusia?
17	tempat	Dimanakah tempat tinggal yang seburuk-buruknya?
18	tempat	Dimanakah tempat bagi yang mendurhakai allah?
19	tempat	Dimanakah letak kiblat umat islam?
20	tempat	Dimanakah tempat yang tidak boleh dihampiri dalam keadaan junub?
21	tempat	Dimanakah orang-orang yang beriman kekal?
22	tempat	Dimanakah matahari terbit?
23	tempat	Dimanakah baitullah dibangun?
24	tempat	Dimanakah harut dan marut diturunkan?
25	tempat	Dimanakah allah bersyiar?
26	tempat	Kemanakah wajahmu dipalingkan?
27	tempat	Kemanakah beribadah haji?
28	tempat	Kemanakah orang yang mendurhakai allah dimasukkan?
29	tempat	Kemanakah kiblat umat islam?
30	tempat	Kemanakah zakariya dipanggil malaikat jibril?
31	tempat	Darimanakah allah menerbitkan matahari?
32	tempat	Darimanakah matahari terbit?
33	tempat	Darimanakah adam diciptakan?
34	tempat	Darimanakah tiga ribu malaikat diturunkan pada perang badar ?
35	tempat	Darimanakah burung diciptakan?
36	waktu	Sampai kapan ibu menyusukan anaknya?
37	waktu	Kapan dihalalkan bercampur dengan isteri selama berpuasa?
38	waktu	Kapankah diturunkan al quran?
39	waktu	Kapankah mengerjakan ibadat haji ?
40	waktu	Kapankah berlaku hukum qishaash?
41	waktu	Kapankah waktu bertasbih yang baik kepada allah?
42	waktu	Kapankah langit dan bumi diciptakan?
43	waktu	Sampai kapankah waktu menyapih?
44	waktu	Sampai kapankah wanita yang ditalak menunggu?
45	waktu	Kapankah isa menjadi saksi?

Masing-masing pertanyaan pada Tabel 5.3 di-*input*-kan ke sistem, kemudian berdasarkan pertanyaan-pertanyaan tersebut pola-pola jawaban yang telah didefinisikan saat membuat sistem akan dinilai berapa banyak pola tersebut memberikan jawaban yang benar dan salah untuk dihitung nilai *confidence* dan *support*-nya. Dari 45 pertanyaan yang telah diuji didapatkan hasil dari masing-masing pola jawaban yang dapat dilihat pada Tabel 5.4.

Tabel 5.4 Hasil Pengujian Aplikasi terhadap Pola Jawaban (*Answer Patterns*)

No	Property	Pola Jawaban	Benar	Salah	Confidence	Support
1	orang	^<PO_orang> <TO>	0	0	0	0
2	orang	<PO_orang> ([a-z])<TO>	1	0	1	0,0023
3	orang	<PO_orang> <TO>	61	73	0,455224	0,1422
4	orang	^apakah <PO_orang><TO>	0	0	0	0
5	orang	^siapa<TO>(.*?)?<PO_orang>	0	0	0	0
6	orang	^<PO_orang> ([a-z])<TO>	0	0	0	0
7	orang	<PO_orang> dalam <TO>	0	0	0	0
8	orang	^<PO_orang> lah <TO>	0	0	0	0
9	orang	wahai <PO_orang> <TO>	2	0	1	0,0047
10	orang	<PO_orang>(.*?)?<TO>	135	462	0,226131	0,3147
11	orang	<TO>(.*?)?<PO_orang>	131	530	0,198185	0,3054
12	orang	<TO> <PO_orang>	20	37	0,350877	0,0466
13	orang	<PO_orang> di waktu<TO>	2	0	1	0,0047
14	orang	ialah <PO_orang> <TO>	4	1	0,8	0,0093
15	orang	apa <TO> <PO_orang>	2	10	0,166667	0,0047
16	orang	ya <PO_orang> <TO>	2	3	0,4	0,0047
17	tempat	<PO_tempat> itu tempat <TO>	3	0	1	0,0171
18	tempat	<TO> ialah<PO_tempat>	1	0	1	0,0057
19	tempat	ialah<PO_tempat>(.*?)?<TO>	1	0	1	0,0057
20	tempat	sungguh <PO_tempat> itu tempat <TO>	1	0	1	0,0057
21	waktu	<TO>(.*?)?<PO_waktu>	24	43	0,358209	0,4000
22	waktu	<PO_waktu>(.*?)?<TO>	10	39	0,204082	0,1667
23	waktu	<TO>(.*?)?di<PO_waktu>	3	10	0,230769	0,0500
24	waktu	^<TO>(.*?)?<PO_waktu>	0	0	0	0
25	orang	adalah <TO><PO_orang>	0	0	0	0
26	orang	^<TO> <PO_orang>	0	0	0	0
27	orang	<TO> itu adalah<PO_orang>	0	0	0	0
28	orang	<PO_orang>(.*?)?karena itu <TO>	0	0	0	0
29	orang	<TO> selain<PO_orang>	6	4	0,6	0,0140
30	orang	<TO> kecuali<PO_orang>	0	0	0	0

Tabel 5.4 Hasil Pengujian Aplikasi terhadap Pola Jawaban (*Answer Patterns*)
(Lanjutan)

No	Property	Pola Jawaban	Benar	Salah	Confidence	Support
31	orang	para <TO> <PO_orang>	0	0	0	0
32	orang	sungguh <PO_orang>(.*)?<TO>	0	0	0	0
33	orang	<PO_orang> itu adalah <TO>	0	0	0	0
34	orang	<PO_orang> hanya <TO>	0	0	0	0
35	orang	^<PO_orang>(.*)?<TO>	0	0	0	0
36	orang	<TO> ialah <PO_orang>	0	0	0	0
37	orang	sungguh <PO_orang> <TO>	0	0	0	0
38	orang	ketika <PO_orang>(.*)?<TO>	1	7	0,125	0,0023
39	orang	seperti <TO> <PO_orang>	0	0	0	0
40	orang	hai <PO_orang> <TO>	2	1	0,666667	0,0047
41	orang	<PO_orang> saja <TO>	0	0	0	0
42	orang	yaitu <PO_orang>(.*)?<TO>	0	0	0	0
43	orang	^<TO>(.*)?<PO_orang>	0	0	0	0
44	orang	^<PO_orang> itu adalah <TO>	0	0	0	0
45	orang	<PO_orang> itu <TO>	1	4	0,2	0,0023
46	tempat	<PO_tempat>(.*)?<TO>	38	126	0,231707	0,2171
47	tempat	<TO>(.*)?di <PO_tempat>	2	11	0,153846	0,0114
48	tempat	itu <PO_tempat> tempat <TO>	1	0	1	0,0057
49	tempat	<TO> <PO_tempat>	7	19	0,269231	0,0400
50	tempat	<TO> (.*)? ke arah <PO_tempat>	0	0	0	0
51	tempat	<PO_tempat> adalah <TO>	1	0	1	0,0057
52	tempat	<TO> ke <PO_tempat>	3	4	0,428571	0,0171
53	tempat	ke <PO_tempat>(.*)?<TO>	0	0	0	0
54	tempat	<TO>(.*)?<PO_tempat>	33	138	0,192982	0,1886
55	tempat	tempat <TO>(.*)?<PO_tempat>	2	0	1	0,0114
56	tempat	<TO> di <PO_tempat>	0	0	0	0
57	tempat	<PO_tempat> dalam <TO>	0	0	0	0
58	tempat	<TO>(.*)?ke dalam <PO_tempat>	4	26	0,133333	0,0229

Tabel 5.4 Hasil Pengujian Aplikasi terhadap Pola Jawaban (*Answer Patterns*)
(Lanjutan)

No	Property	Pola Jawaban	Benar	Salah	Confidence	Support
59	tempat	tempat <TO> <PO_tempat>\$	0	0	0	0
60	waktu	<PO_waktu>(.*)?dalam <TO>	1	0	1	0,0167
61	waktu	<TO> sudah <PO_waktu>	0	0	0	0
62	waktu	masa <PO_waktu>(.*)?<TO>	0	0	0	0
63	waktu	yaitu <PO_waktu>(.*)?<TO>	0	0	0	0
64	waktu	<TO> <PO_waktu> lama	0	0	0	0
65	waktu	<TO> <PO_waktu>	8	16	0,333333	0,1333
66	waktu	<TO>(.*)?lama <PO_waktu>	1	0	1	0,0167
67	waktu	<TO> di waktu <PO_waktu> hari	0	0	0	0
68	waktu	<PO_waktu> saja <TO>	0	0	0	0
69	waktu	di <PO_waktu>(.*)?<TO>	4	7	0,363636	0,0667

Tabel 5.4 di atas menunjukkan jumlah jawaban yang benar yang diberikan oleh pola jawaban, jumlah jawaban yang salah yang diberikan oleh pola jawaban, nilai *confidence* dan *support* dari masing-masing pola yang dihasilkan dari perhitungan sesuai dengan persamaan 2.4 dan 2.5. Penjelasan lebih rinci mengenai Tabel 5.4 sebagai berikut.

a. Orang

- 1) Pola ^<PO_orang> <TO> menghasilkan 0 buah jawaban yang benar dan 0 buah jawaban yang salah dari 15 pertanyaan yang di-*input*-kan untuk kategori orang dengan nilai *confidence* sebesar 0 dan nilai *support* sebesar 0.
- 2) Pola $\text{<PO_orang> ([a-z]) <TO>}$ menghasilkan 1 buah jawaban yang benar dan 0 buah jawaban yang salah dari 15 pertanyaan yang di-*input*-kan untuk kategori orang dengan nilai *confidence* sebesar 1 dan nilai *support* sebesar 0,0023.

- 3) Pola <PO_orang> <TO> menghasilkan 61 buah jawaban yang benar dan 73 buah jawaban yang salah dari 15 pertanyaan yang di-input-kan untuk kategori orang dengan nilai *confidence* sebesar 0,455224 dan nilai *support* sebesar 0,1422.
- 4) Pola ^apakah <PO_orang> <TO> menghasilkan 0 buah jawaban yang benar dan 0 buah jawaban yang salah dari 15 pertanyaan yang di-input-kan untuk kategori orang dengan nilai *confidence* sebesar 0 dan nilai *support* sebesar 0.
- 5) Pola ^siapa <TO>(.*?)<PO_orang> menghasilkan 0 buah jawaban yang benar dan 0 buah jawaban yang salah dari 15 pertanyaan yang di-input-kan untuk kategori orang dengan nilai *confidence* sebesar 0 dan nilai *support* sebesar 0.
- 6) dan seterusnya.

b. Tempat

- 1) Pola <PO_tempat> itu tempat <TO> menghasilkan 3 buah jawaban yang benar dan 0 buah jawaban yang salah dari 20 pertanyaan yang di-input-kan untuk kategori tempat dengan nilai *confidence* sebesar 1 dan nilai *support* sebesar 0,0171.
- 2) Pola <TO> ialah <PO_tempat> menghasilkan 1 buah jawaban yang benar dan 0 buah jawaban yang salah dari 20 pertanyaan yang di-input-kan untuk kategori tempat dengan nilai *confidence* sebesar 1 dan nilai *support* sebesar 0,0057.
- 3) Pola ialah <PO_tempat>(.*?)<TO> menghasilkan 1 buah jawaban yang benar dan 0 buah jawaban yang salah dari 20 pertanyaan yang di-input-kan untuk kategori tempat dengan nilai *confidence* sebesar 1 dan nilai *support* sebesar 0,0057.
- 4) Pola sungguh <PO_tempat> itu tempat <TO> menghasilkan 1 buah jawaban yang benar dan 0 buah jawaban yang salah dari 20

pertanyaan yang di-*input*-kan untuk kategori tempat dengan nilai *confidence* sebesar 1 dan nilai *support* sebesar 0,0057.

5) Pola <PO_tempat>(.*)?<TO> menghasilkan 38 buah jawaban yang benar dan 126 buah jawaban yang salah dari 20 pertanyaan yang di-*input*-kan untuk kategori tempat dengan nilai *confidence* sebesar 0,231707 dan nilai *support* sebesar 0,2171.

6) dan seterusnya.

c. Waktu

1) Pola <TO>(.*)?<PO_waktu> menghasilkan 24 buah jawaban yang benar dan 43 buah jawaban yang salah dari 10 pertanyaan yang di-*input*-kan untuk kategori waktu dengan nilai *confidence* sebesar 0,358209 dan nilai *support* sebesar 0,4000.

2) Pola <PO_waktu>(.*)?<TO> menghasilkan 10 buah jawaban yang benar dan 39 buah jawaban yang salah dari 10 pertanyaan yang di-*input*-kan untuk kategori waktu dengan nilai *confidence* sebesar 0,204082 dan nilai *support* sebesar 0,1667.

3) Pola <TO>(.*)?di <PO_waktu> menghasilkan 3 buah jawaban yang benar dan 10 buah jawaban yang salah dari 10 pertanyaan yang di-*input*-kan untuk kategori waktu dengan nilai *confidence* sebesar 0,230769 dan nilai *support* sebesar 0,0500.

4) Pola ^<TO>(.*)?<PO_waktu> menghasilkan 0 buah jawaban yang benar dan 0 buah jawaban yang salah dari 10 pertanyaan yang di-*input*-kan untuk kategori waktu dengan nilai *confidence* sebesar 0 dan nilai *support* sebesar 0.

5) Pola <PO_waktu>(.*)?dalam <TO> menghasilkan 1 buah jawaban yang benar dan 0 buah jawaban yang salah dari 10 pertanyaan yang di-*input*-kan untuk kategori waktu dengan nilai *confidence* sebesar 1 dan nilai *support* sebesar 0,0167.

6) dan seterusnya.

Pada penelitian ini jika nilai *confidence* dan *support*-nya 0, maka pola jawaban tersebut akan dibuang guna mempercepat komputasi. Dari Tabel 5.4 maka pola jawaban yang tidak digunakan untuk mengekstrak jawaban adalah pola jawaban nomor 1, 4, 5, 6, 8, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 39, 41, 42, 43, 44, 50, 53, 56, 57, 59, 61, 62, 63, 64, 67 dan 68.

5.2.2.2 Pengujian terhadap Aspek Presisi (*Precision*) dan CWS

Dalam mengevaluasi aplikasi *QA System* ini dilakukan dua kali pengujian. Pengujian pertama adalah pengujian yang penulis lakukan dengan memberikan pertanyaan-pertanyaan yang sesuai dengan pola-pola pertanyaan yang telah didefinisikan pada sistem. Pada pengujian pertama ini, dimasukkan 30 pertanyaan, dimana pertanyaan-pertanyaan tersebut merupakan pertanyaan baru yang belum diajukan pada pengujian pola jawaban. Pertanyaan-pertanyaan baru tersebut terdiri dari 10 pertanyaan dengan kategori orang, 10 pertanyaan dengan kategori waktu, dan 10 pertanyaan dengan kategori tempat. Dalam pengujian ini diasumsikan bahwa tiap-tiap pertanyaan memiliki minimal satu jawaban. Hasil pengujian ini dapat dilihat pada Tabel 5.5.

Pengujian kedua adalah pengujian yang dilakukan oleh pengguna umum yang pada pengujian ini pengguna tersebut memiliki pemahaman yang baik mengenai Al Quran dan tidak mengetahui pola/aturan yang diberikan pada aplikasi ini. Pada pengujian ini pengguna memasukkan 10 pertanyaan sesuai dengan kategori yang telah ditentukan. Hasil pengujian ini dapat dilihat pada Tabel 5.6.

Dalam menilai apakah jawaban yang dikembalikan oleh sistem bernilai benar atau tidak, ditetapkan sejumlah aturan sebagai berikut.

- 1) sebuah jawaban yang diberikan sistem dari pertanyaan yang diajukan *user* akan bernilai benar, jika:
 - a) jawaban dan dokumen yang dikembalikan adalah benar (*correct*).

- b) jawaban dan dokumen yang dikembalikan benar, namun terlalu panjang (*not exact*).
- 2) sebuah jawaban yang diberikan sistem dari pertanyaan yang diajukan *user* akan bernilai salah, jika:
 - a) jawaban tidak benar atau jawabannya tidak responsif (*incorrect*);
 - b) jawaban benar namun dokumen yang dikembalikan tidak mendukung jawaban tersebut (*not supported*).

Tabel 5.5 dan 5.6 menjelaskan mengenai nilai *precision of correct* yang didapat dari persamaan 2.10, *precision of incorrect* yang didapat dari persamaan 2.11 dan ranking posisi dari jawaban benar yang pertama kali ditemukan dimana jawaban yang benar tersebut tergolong ada kategori *correct* dan *not exact*. Ranking posisi digunakan untuk menghitung nilai CWS dari hasil pengujian sesuai persamaan 2.12.

Tabel 5.5 Hasil Pengujian Aplikasi Terhadap Aspek Presisi dan CWS pada Pengujian Pertama

No	Pertanyaan	Kategori	Jawaban yang diharapkan	Jumlah jawaban yang ditemukan	Jumlah				Rangking (urutan)*	PC* (%)	PI* (%)
					Benar		Salah				
					C*	NE*	IC*	NS*			
1	Diwahyukan kepada siapakah al quran?	orang	Muham-mad, rasul	69	4	0	63	2	28	5,797	94,203
2	Kepada siapakah iblis tidak sujud?	orang	adam, manusia	22	7	0	15	0	6	31,818	68,182
3	Siapakah yang sujud kepada adam?	orang	malaikat	19	5	0	14	0	11	26,316	73,684
4	Siapakah yang mengharamkan bangkai, darah, daging babi, dan binatang yang (ketika disembelih) disebut (nama) selain Allah?	orang	allah	22	6	0	7	9	6	27,273	72,727
5	Siapakah yang melawan jalut dan tentaranya?	orang	thabut	4	1	0	3	0	4	25	75
6	Siapakah yang menggoda yusuf?	orang	zulaikha, wanita-wanita	6	2	0	4	0	5	33,333	66,667
7	Siapakah putera adam?	orang	habil, qabil	13	2	0	11	0	8	15,385	84,615
8	Siapakah yang diciptakan dari nyala api?	orang	jin	10	3	0	7	0	3	30	70

Keterangan:

C : *Correct*

NE : *Not Exact*

IC : *Incorrect*

NS : *Not Support*

PC : *Precision of Correct*

PI : *Precision of Incorrect*

Rangking (urutan) : Posisi jawaban benar yang ditemukan pertama kali

Tabel 5.5 Hasil Pengujian Aplikasi Terhadap Aspek Presisi dan CWS pada Pengujian Pertama (Lanjutan)

No	Pertanyaan	Kategori	Jawaban yang diharapkan	Jumlah jawaban yang ditemukan	Jumlah				Rangking (urutan)*	PC* (%)	PI* (%)
					Benar		Salah				
					C*	NE*	IC*	NS*			
9	Siapakah yang menyembelih anak laki-laki dan membiarkan anak perempuan hidup?	orang	firaun	18	3	0	15	0	14	16,667	83,333
10	Siapakah yang menurunkan air dari langit?	orang	Allah	31	10	0	19	2	3	32,258	67,742
11	Darimanakah burung diciptakan?	tempat	Tanah	3	2	0	1	0	2	66,667	33,333
12	Dimanakah harut dan marut diturunkan?	tempat	Babil	1	1	0	0	0	1	100	0
13	Dimanakah letak makam Ibrahim?	tempat	Baitullah	1	1	0	0	0	1	100	0
14	Darimanakah air hujan turun?	tempat	langit	57	23	0	34	0	3	40,351	59,649
15	Kemanakah musa dilemparkan?	tempat	Sungai nil	3	2	0	1	0	2	66,667	33,333
16	Dimanakah muhammad melihat jibril dalam rupanya yang asli?	tempat	sidratil muntaha	1	1	0	0	0	1	100	0

Keterangan:

C : *Correct*NE : *Not Exact*IC : *Incorrect*NS : *Not Support*PC : *Precision of Correct*PI : *Precision of Incorrect*

Rangking (urutan) : Posisi jawaban benar yang ditemukan pertama kali

Tabel 5.5 Hasil Pengujian Aplikasi Terhadap Aspek Presisi dan CWS pada Pengujian Pertama (Lanjutan)

No	Pertanyaan	Kategori	Jawaban yang diharapkan	Jumlah jawaban yang ditemukan	Jumlah				Rangking (urutan)*	PC* (%)	PI* (%)
					Benar		Salah				
					C*	NE*	IC*	NS*			
17	Dimanakah hud memberi peringatan kepada kaum 'aad?	tempat	Al ahqaaf	6	1	0	5	0	2	16,667	83,333
18	Dimanakah tempat yang warnanya hijau tua?	tempat	surga	1	1	0	0	0	1	100	0
19	Dimanakah musa mendapati dua orang laki-laki yang berkelahi?	tempat	Kota memphis	2	2	0	0	0	1	100	0
20	Dimanakah tempat bertolaknya orang berhaji?	tempat	Arafah	5	1	0	4	0	1	20	80
21	Sampai kapankah diwajibkan berpuasa bagi orang yang tidak menyembelih korban pada masa haji?	waktu	Tiga hari	4	1	0	3	0	1	25	75
22	Sampai kapankah wanita yang ditalak menunggu?	waktu	Tiga kali quru'	1	1	0	0	0	1	100	0
23	Kapankah disempurnakannya pahala?	waktu	Hari kiamat	5	1	0	4	0	5	20	80
24	Sampai kapankah isteri diilaa?	waktu	Empat bulan	8	0	0	8	0	0	0	100

Keterangan:

C : *Correct*

NE : *Not Exact*

IC : *Incorrect*

NS : *Not Support*

PC : *Precision of Correct*

PI : *Precision of Incorrect*

Rangking (urutan) : Posisi jawaban benar yang ditemukan pertama kali

Tabel 5.5 Hasil Pengujian Aplikasi Terhadap Aspek Presisi dan CWS pada Pengujian Pertama (Lanjutan)

No	Pertanyaan	Kategori	Jawaban yang diharapkan	Jumlah jawaban yang ditemukan	Jumlah				Rangking (urutan)*	PC* (%)	PI* (%)
					Benar		Salah				
					C*	NE*	IC*	NS*			
25	Kapankah harta yang dibakhilkan akan dikalungkan ke lehernya?	waktu	Hari kiamat	6	1	0	5	0	1	16,667	83,333
26	Kapankah berdoa memohon ampun yang baik?	waktu	Sahur	6	1	0	5	0	1	16,667	83,333
27	Kapan taurat diberikan?	waktu	Empat puluh malam	7	2	0	5	0	1	28,571	71,429
28	Kapan kamu sekalian dikumpulkan allah?	waktu	Hari kiamat, hari akhirat, hari pengumpulan, hari kemudian , hari bangkit, hari balas	73	3	0	57	13	12	4,110	95,890
29	Kapankah isa menjadi saksi?	waktu	Hari kiamat	1	1	0	0	0	1	100	0

Keterangan:

C : *Correct*NE : *Not Exact*IC : *Incorrect*NS : *Not Support*PC : *Precision of Correct*PI : *Precision of Incorrect*

Rangking (urutan) : Posisi jawaban benar yang ditemukan pertama kali

Tabel 5.5 Hasil Pengujian Aplikasi Terhadap Aspek Presisi dan CWS pada Pengujian Pertama (Lanjutan)

No	Pertanyaan	Kategori	Jawaban yang diharapkan	Jumlah jawaban yang ditemukan	Jumlah				Rangking (urutan)*	PC* (%)	PI* (%)
					Benar		Salah				
					C*	NE*	IC*	NS*			
30	Sampai kapankah bumi diciptakan?	waktu	Dua masa	25	1	0	24	0	13	4	96

Keterangan:

C : *Correct*

NE : *Not Exact*

IC : *Incorrect*

NS : *Not Support*

PC : *Precision of Correct*

PI : *Precision of Incorrect*

Rangking (urutan) : Posisi jawaban benar yang ditemukan pertama kali

Tabel 5.6 Hasil Pengujian Aplikasi Terhadap Aspek Presisi dan CWS pada Pengujian Kedua

No	Pertanyaan	Kategori	Jawaban yang diharapkan	Jumlah jawaban yang ditemukan	Jumlah				Rangking (urutan)*	PC* (%)	PI* (%)
					Benar		Salah				
					C*	NE*	IC*	NS*			
1	Siapakah yang menurunkan air hujan?	orang	Allah, tuhan	29	13	0	15	1	1	44,828	55,172
2	Siapakah yang dibunuh oleh nabi daud?	orang	Jalut	35	1	0	34	0	6	2,857	97,143
3	Kapankah diturunkannya al quran?	waktu	Bulan Ramadhan, malam mulia, malam berkah	18	5	0	13	0	1	27,778	72,222
4	Siapakah yang membuat perumpamaan berupa nyamuk?	orang	Allah	25	1	0	16	8	25	4	96
5	Siapakah yang dianugerahi ilmu yang luas dan tubuh yang perkasa?	orang	Raja Thalut	4	2	0	2	0	1	50	50
6	Siapakah yang memukulkan tongkatnya ke atas batu?	orang	Musa	0	0	0	0	0	0	0	0
7	Dimanakah letak masjidil haram?	tempat	Mekah	4	3	0	1	0	2	75	25

Keterangan:

C : *Correct*

NE : *Not Exact*

IC : *Incorrect*

NS : *Not Support*

PC : *Precision of Correct*

PI : *Precision of Incorrect*

Rangking (urutan) : Posisi jawaban benar yang ditemukan pertama kali

Tabel 5.6 Hasil Pengujian Aplikasi Terhadap Aspek Presisi dan CWS pada Pengujian Kedua (Lanjutan)

No	Pertanyaan	Kategori	Jawaban yang diharapkan	Jumlah jawaban yang ditemukan	Jumlah				Rangking (urutan)*	PC* (%)	PI* (%)
					Benar		Salah				
					C*	NE*	IC*	NS*			
8	Siapakah yang diusir Allah dari surga?	orang	Adam	20	0	0	0	0	0	0	0
9	Siapakah yang menjadi khalifah di muka bumi?	orang	Manusia, Daud	12	2	0	9	1	1	16,667	83,333
10	Siapakah yang menghuni neraka kelak?	orang	Orang-orang kafir, orang zalim, orang musyrik, syaitan	33	12	0	21	0	2	36,364	63,636

Keterangan:

C : *Correct*

NE : *Not Exact*

IC : *Incorrect*

NS : *Not Support*

PC : *Precision of Correct*

PI : *Precision of Incorrect*

Rangking (urutan) : Posisi jawaban benar yang ditemukan pertama kali

Berdasarkan Tabel 5.5 dan 5.6 di atas nilai presisi dan CWS dapat dilihat pada Tabel 5.7.

Tabel 5.7 Nilai Presisi dan CWS

No	Pengujian	Jumlah pertanyaan	Rata-rata <i>Precision of Correct (%)</i>	Rata-rata <i>Precision of Incorrect (%)</i>	CWS
1	Pengujian pertama	30	42,31	57,69	0,011
2	Pengujian kedua (pengguna umum)	10	25,75	54,25	0,022

5.2.2.3 Pengujian berdasarkan Interpretasi Pertanyaan

Pertanyaan-pertanyaan yang diujikan pada sub bab ini disesuaikan dengan pola pertanyaan yang telah diberikan kepada sistem. Tabel 5.8 memperlihatkan hasil pengujian.

Tabel 5.8 Hasil Pengujian berdasarkan Interpretasi Pertanyaan

No	Objek Pengujian	Pertanyaan	Interpretasi Sistem	Jawaban yang diharapkan	Jumlah Dokumen jawaban yang ditemukan	Jumlah Dokumen Jawaban yang benar	Posisi Jawaban yang ditemukan benar pertama kali (urutan)
1	Menanyakan orang yang menerima wahyu al quran	Kepada siapakah al quran diwahyukan?	<i>Property:</i> Orang <i>Target:</i> quran <i>Context:</i> wahyu	Muhammad, rasul	291	19	6
		Al quran diwahyukan kepada siapa?	<i>Property:</i> Orang <i>Target:</i> wahyu <i>Context:</i> quran		71	4	7
		Diwahyukan kepada siapakah al quran?	<i>Property:</i> Orang <i>Target:</i> wahyu <i>Context:</i> quran		69	3	26
2	Menanyakan letak baitullah	Dimanakah letak baitullah?	<i>Property:</i> Tempat <i>Target:</i> baitullah <i>Context:</i> -	Mekah, Bakkah	12	2	10
		Dimanakah Baitullah berada?	<i>Property:</i> Tempat <i>Target:</i> baitullah <i>Context:</i> -		0	0	0
		Dimanakah Baitullah dibangun?	<i>Property:</i> Tempat <i>Target:</i> bangun <i>Context:</i> baitullah	Mekah, Bakkah	9	2	1

Tabel 5.8 Hasil Pengujian berdasarkan Interpretasi Pertanyaan (Lanjutan)

No	Objek Pengujian	Pertanyaan	Interpretasi Sistem	Jawaban yang diharapkan	Jumlah Dokumen jawaban yang ditemukan	Jumlah Dokumen Jawaban yang benar	Posisi Jawaban yang ditemukan benar pertama kali (urutan)
3	Menanyakan tempat yang paling buruk (neraka)	Dimanakah tempat tinggal yang seburuk-buruknya?	<i>Property:</i> Tempat <i>Target:</i> tinggal <i>Context:</i> buruk buruk	Neraka, jahanam	9	8	1
		Dimanakah tempat seburuk-buruk kembali?	<i>Property:</i> Tempat <i>Target:</i> buruk <i>Context:</i> kembali		22	21	1
		Dimanakah seburuk-buruknya tempat kembali?	<i>Property:</i> Tempat <i>Target:</i> kembali <i>Context:</i> buruk buruk tempat		17	16	1
4	Menanyakan kapan turunnya al quran	Kapan diturunkannya al quran?	<i>Property:</i> Waktu <i>Target:</i> turun <i>Context:</i> quran	Bulan Ramadhan	18	3	1
		Kapan al quran diturunkan?	<i>Property:</i> Waktu <i>Target:</i> quran <i>Context:</i> turun		20	1	11

Tabel 5.8 Hasil Pengujian berdasarkan Interpretasi Pertanyaan (Lanjutan)

No	Objek Pengujian	Pertanyaan	Interpretasi Sistem	Jawaban yang diharapkan	Jumlah Dokumen jawaban yang ditemukan	Jumlah Dokumen Jawaban yang benar	Posisi Jawaban yang ditemukan benar pertama kali (urutan)
5	Menanyakan orang yang menjadi saksi bagi ahli kitab di hari kiamat	Siapakah yang menjadi saksi bagi ahli kitab pada hari kiamat?	<i>Property:</i> Orang <i>Target:</i> saksi <i>Context:</i> ahli kitab hari kiamat	Isa	39	1	4
		Siapakah yang menjadi saksi pada hari kiamat?	<i>Property:</i> Orang <i>Target:</i> saksi <i>Context:</i> hari kiamat		22	1	4

5.2.3 Kesimpulan Pengujian

Dari pengujian-pengujian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Berdasarkan pengujian pertama yaitu pengujian yang penulis lakukan pada pengujian maka aplikasi *QA System* ini mempunyai nilai *precision of correct* sebesar 42,31%, *precision of incorrect* 57,69% dan CWS sebesar 0,011 pada 30 buah pertanyaan (Lihat Tabel 5.7).
2. Berdasarkan pengujian kedua yaitu pengujian yang diberikan oleh pengguna umum pada pengujian maka aplikasi *QA System* ini mempunyai nilai *precision of correct* sebesar 25,75%, *precision of incorrect* 54,25% dan CWS sebesar 0,022 pada 10 buah pertanyaan (Lihat Tabel 5.7).
3. Interpretasi terhadap pertanyaan dapat mempengaruhi kinerja dari sistem seperti jumlah jawaban yang ditemukan, jumlah jawaban yang benar dan ranking jawaban (Lihat Tabel 5.8).

BAB VI

PENUTUP

6.1. Kesimpulan

Berdasarkan dari penelitian yang telah dilakukan maka dapat ditarik kesimpulan sebagai berikut:

1. Secara umum, *Ephyra framework* dapat diterapkan untuk rancang bangun *Question Answering System* pada terjemahan Al Quran berbasis web dengan kategori pertanyaan yaitu orang, waktu, dan tempat. Namun secara fungsional, aplikasi ini tidak direkomendasikan karena tingkat presisi jawaban aplikasi yang rendah sehingga akan berpengaruh buruk terhadap persepsi pengguna dalam memahami Al Quran.
2. Aplikasi *QA System* dengan menggunakan *Ephyra framework* ini belum efektif terhadap bahasa Indonesia hasil terjemahan dari bahasa asing pada terjemahan Al Quran.
3. Hasil aplikasi berupa jawaban singkat dan potongan dokumen yang berisi jawaban dari pertanyaan yang diajukan.
4. Penggunaan *regular expression* pada pola (*pattern*) menjadikan pola lebih fleksibel, efektif dan mudah diterapkan.
5. Semakin spesifik/unik pertanyaan yang diajukan maka semakin baik jawaban yang diberikan oleh sistem.
6. Ketepatan interpretasi sistem terhadap pertanyaan mempengaruhi keakuratan jawaban dari pertanyaan tersebut.
7. Performansi dari aplikasi *Question Answering System* ini dipengaruhi oleh beberapa faktor yaitu:

- a. Bahasa Indonesia pada terjemahan Al Quran merupakan hasil terjemahan dari bahasa asing dan banyak menggunakan kata kiasan atau perumpamaan.
- b. Penyajian dan penentuan sumber informasi yang digunakan untuk menjawab pertanyaan yaitu korpus dan basis data *stoplist*.
- c. Penyajian dan penentuan pola pertanyaan, pola jawaban, dan kamus jawaban yang menjadi sumber pengetahuan sistem.
- d. Pemilihan teknik *stemming*, pemodelan IR dan penyeleksian jawaban.
- e. Sasaran *user* yang menjadi tujuan penggunaan aplikasi. Mengingat bahwa Al Quran merupakan pedoman wajib yang harus diikuti oleh umat Islam maka informasi-informasi harus disajikan secara benar.
- f. Tipe Pertanyaan.

6.2. Saran

Adapun saran-saran yang diajukan oleh penulis mengenai penelitian ini adalah sebagai berikut:

1. Pada pengembangan selanjutnya diharapkan aplikasi QA System tidak hanya terbatas pada pertanyaan dengan kategori orang, waktu, dan tempat saja namun dapat pula menjawab pertanyaan dengan kategori yang lainnya.
2. Mengembangkan aplikasi dengan tiga jenis ekstraksi jawaban yang dimiliki Ephyra yaitu *pattern learning and matching*, *semantic parsing*, dan *answer type analysis* agar aplikasi memiliki nilai presisi dan relevansi yang tinggi.
3. Pada pengembangan selanjutnya aplikasi tidak mendefinisikan kamus jawaban secara manual tetapi dapat menggunakan teknik-teknik yang dapat mengenali jawaban secara otomatis.
4. Pengindeksan dan segmentasi kalimat dapat dilakukan secara otomatis yang tidak membutuhkan waktu komputasi yang lebih lama, sehingga adanya

penambahan ataupun pengurangan dokumen tidak perlu dilakukan pengindeksan dan segmentasi ulang.

DAFTAR PUSTAKA

- Bunyamin, Hendra dan Negara, Chathalea Puspa. *Aplikasi Information Retrieval (IR) CATA Dengan Metode Generalized Vector Space Model*. Jurnal Informatika, vol. 4, no. 1, pp. 29-38. 2008.
- Chaer, Abdul. *Sintaksis Bahasa Indonesia: Pendekatan Proses*. PT Rineka Cipta, Jakarta. 2009.
- Dang, H. Trang, dkk. *Overview of The TREC 2006 Question Answering TRACK*. Proceedings of the Fifteenth Text Retrieval Conference. 2007.
- Fornier, Pamela, dkk. *Overview of The CLEF 2008 Multilingual Question Answering TRACK*. 2008.
- Gunawan., Purwanto, Devi Dwi, dkk. *Penyusunan Strong's Concordance untuk Alkitab Perjanjian Baru Bahasa Indonesia*. Jurnal Sistem Informasi MTI UI, vol. 5, no. 2, ISBN 1412-8896. 2012.
- Gunawan dan Lovina, Gita. *Question Answering System dan Penerapannya pada Alkitab*. Jurnal Informatika, vol. 7, no. 1, pp. 1-9. 2006
- Mandala, Rila. *Evaluasi Kinerja Sistem Penyaringan Informasi Model Ruang Vektor*. Seminar Nasional Aplikasi Teknologi Informasi 2006 (SNATI 2006), ISSN 1907-5022. 2006
- Nurdiati, Sri, dkk. *Query-Sensitive Similarity Measure dalam Temu Kembali Dokumen Berbahasa Indonesia*. Bogor: Fakultas Matematika dan IPA, Institut Pertanian Bogor. 2010.
- Schlaefter, Nico. *Pattern Learning and Knowledge Annotation for Question Answering*. USA: Carnegie Mellon University. 2005.
- Schlaefter, Nico. *Statistical Source Expansion for Question Answering*. Diploma Thesis. USA: Language Technologies Institute School of Computer Science Carnegie Mellon University. 2009-2011.

Schlaefter, Nico., Gieselmann, Petra dan Sautter, G. *The Ephyra QA System at TREC 2006*. Proceedings of the Fifteenth Text Retrieval Conference. 2006.

Schlaefter, Nico., Gieselmann, Petra, dkk. *A Pattern Learning Approach to Question Answering within the Ephyra Framework*. Springer-Verlag Berlin Heidelberg, LNAI 4188, pp. 687-694. 2006

Toba, Hapnes dan Adriani, Mirna. *Pattern Based Approach in Indonesian Question-Answering System*. Jakarta: Information Retrieval Laboratory Fakultas Ilmu Komputer, Universitas Indonesia. 2009.

_____. "Website Ask Jeeves". [Online] available. www.ask.com, 4 November 2012.

_____. "Website Ephyra". [Online] available. www.ephyra.info, 4 November 2012.

_____. "Website Yahoo Answer". [Online] available. id.answers.yahoo.com, 4 November 2012.

Zhang, D dan Lee, W. *Web Based Pattern Mining and Matching Approach to Question Answering*. Proceedings of the 11th Text Retrieval Conference (TREC). 2002.

Zheng, Z. *AnswerBus Question Answering System*. USA: University of Michigan. 2002.